THE DIVERSITY, DISTRIBUTION AND POTENTIAL METABOLISM OF NON-CYANOBACTERIAL DIAZOTROPHS IN THE NORTH ATLANTIC OCEAN

by

Jenni-Marie Ratten

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Dalhousie University

Halifax, Nova Scotia

June 2017

~ Michael Leunig

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In large areas of the Atlantic Ocean, primary productivity is limited by fixed nitrogen. Biological N_2 fixation, conducted by diazotrophs, contributes significantly to the input of new fixed nitrogen in those areas. Recent findings clearly showed the occurrence of non-cyanobacterial diazotrophs in oceanic areas outside of the most frequently investigated cyanobacteria-dominated tropical surface oceans. This thesis investigates the diazotrophic community in the coastal and open Atlantic Ocean spanning from surface to depth and from 40°S to 60°N.

Using TaqMan assays, the abundances of several diazotrophic phylotypes were measured in the context of the US GEOTRACES program on a transect in the tropical Atlantic Ocean. Distribution of diazotrophs in the surface waters was significantly correlated with the deposition of Saharan dust in the Eastern North Atlantic. Below the surface, an association with the nutrient-rich North African upwelling waters was found.

High-throughput sequencing of the *nifH* gene from 407 samples collected throughout the Atlantic, uncovered a broad array of new *nifH* sequences. It further demonstrated the shift from cyanobacterial-dominated diazotrophic communities in the tropics to non-cyanobacterial communities at higher latitudes and below the euphotic zone.

To better understand the functional role of these marine diazotrophs, which have few ecologically relevant representative in culture, an analysis of 132 diazotrophic reference genomes including 112 non-cyanobacterial species revealed their diverse metabolic potential. Utilization of alternative organic carbon sources, iron acquisition and anaerobic respiration were some aspects found, indicating a role that exceeds that of N_2 fixation.

A novel heterotrophic diazotroph was isolated from the Bedford Basin using cell-sorting flow cytometry. The isolate could grow in artificial seawater depleted in fixed nitrogen and was actively transcribing the *nifH* gene. Genome sequencing revealed the presence of the full *nif* operon and pathways that suit diazotrophy. Clade-specific TaqMan qPCR assay showed the wide distribution of the isolate in the temperate North Atlantic Ocean.

A one-year time-series explored the entire microbial community and the diazotrophs in the temperate Bedford Basin, where the novel diazotroph was isolated. Temperature, nutrients and O_2 concentrations were the major drivers of microbial community structure. The diazotrophic community was very diverse and showed seasonal variation.

LIST OF ABBREVIATIONS AND SYMBOLS USED

16S rRNA	16S ribosomal RNA gene
α	alpha
Al	Aluminium
Anammox	Anaerobic Ammonia Oxidation
ANOSIM	Analysis of Similarities
AZMP	Atlantic Zone Monitoring Programme
β	beta
BATS	Bermuda Time Series
BCO-DMO	Biological and Chemical Oceanography Data Management Office
BEST	Bio-Env + Stepwise
BIO	Bedford Institute of Oceanography
BNF	Biological nitrogen fixation
BLAST	Basic Local Alignment Tool
bp	base pair
CGEB	Centre for Comparative and Evolutionary Biology
Cr	Chromium
CRISPR/Cas	Clustered regularly interspaced short palindromic repeats/CRISPR-associated system

CTD	Instrument used to measure Conductivity, Temperature and Depth in oceanographic context
CVOO	Cape Verde Ocean Observatory
δ	delta
DIN	Dissolved Inorganic Nitrogen
DIP	Dissolved Organic Phosphate
DNA	Deoxyribonucleic acid
ε	epsilon
FACS	Fluorescent-Activated Cell Sorting
Fe	Iron
FISH	Fluorescent In Situ Hybridisation
γ	gamma
Gamma A	γ-Proteobacterium A
Het1	*Rhizosolenia-Richelia*
Het2	*Hemiaulus-Richelia*
Hg	Mercury
HOT	Hawaii Ocean Time-series
IMR	Integrated Microbiome Resource
iTOL	Interactive Tree of Life
LINKTREE	Linkage tree

LGT	Lateral Gene Transfer
µ	micro
MAR	Mid-Atlantic Ridge
MLD	Mixed Layer Depth
Mg	Magnesium
N_2	Dinitrogen gas
Na	Sodium
Ni	Nickel
NMDS	Non-linear Multi-Dimensional Scaling
O_2	Oxygen
ODV	Ocean Data View
OMZ	Oxygen Minimum Zone
OTU	Operational Taxonomic Unit
PCA	Principle component analysis
PCR	Polymerase Chain Reaction
PEAR	Paired-End reAd merger
PHA	Polyhydroxyalkanoates
PSU	Practical Salinity Unit
PyNAST	Python Nearest Alignment Space Termination

QIIME	Quantitative Insights Into Microbial Ecology
qPCR	quantitive Polymerase Chain Reaction
RAxML	Randomized Axelerated Maximum Likelihood
RNA	Ribonucleic acid
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RuMP	Ribulose Monophosphate Pathway
Se	Selenium
SMBO	Santa Monica Bay Observatory
SML	Surface Mixed Layer
SPOTS	San Pedro Ocean Time Series
UCYN A	Unicellular Cyanobacterium Group A; *Candidatus* Atelocyanobacterium thalassa
UCYN B	Unicellular Cyanobacterium Group B; *Crocosphaera*
UCYN C	Unicellular Cyanobacterium Group C; *Cyanothece*
V6-V8	Variable region 6 to Variable region 8 of the 16S rRNA gene
Zn	Zinc

Units and prefixes were used according to the International System of Units with the exceptions of molecular concentrations and rotations per minute, for which M instead of mol/l and rpm instead of 1/min were used.

ACKNOWLEDGEMENTS

My first thank you goes to my supervisor, Dr. Julie LaRoche, for dedicating an interesting research topic to me and for the possibility of completing my PhD project in her research group. I am also very thankful for the support throughout the years, for listening to my struggles and successes as well as providing food for thought when needed. She made the work environment positive and I thoroughly enjoyed working in that atmosphere.

I would like to thank my co-supervisor, Dr. Ruth Schmitz-Streit, who provided me with her resources, advice and support during my research stays in Germany and abroad when in Canada without hesitation. She made me feel very welcome in her research group and I am still in productive exchanges with many people she introduced me to.

Thank you to my committee members, Dr. Alastair Simpson and Dr. Morgan Langille, who guided me smoothly though my research. They provided me with constructive feedback, solutions and resources from their fields on countless occasions that continuously drove my work forward.

A big thank you to Dr. Dhwani Desai for introducing me to and guiding me through the world of bioinformatics and statistics. Without his help, great parts of this thesis would not have been possible. Thank you for your patience and for finding all those missing semicolons.

An indispensable and professional help in and outside the laboratory was Dr. Jennifer Tolman. Her technical support as well as ideas and feedback on manuscripts made my work much easier.

Thank you to my other laboratory partners: Dr. Rebecca Langlois, Jörg Behnke, Jaqueline Zorz, Ian Luddington, Dr. Stephanie Michl, Hannah Blanchard, Jude van der Meer and Sallie Lau who provided excellent professional and social support to me. Constructive discussions, ideas and suggestions allowed my work to progress into directions that would have not happened without them. I very much appreciated their good company.

I would also like to thank Dr. Andre Comeau and the rest of the IMR (Integrated Microbiome Resource) team for performing all of my high-throughput sequencing work and for taking the time and effort to assist me through the analysis of the data.

Without TOSST (Transatlantic Ocean Systems Science and Technology), my PhD research would not have been the same. The group of students as well as Dr. Douglas Wallace and Dr. Markus Kienast provided extra support, feedback, ideas and opportunities that extended my experience at Dalhousie University beyond the subject of biology. The interdisciplinary cooperation including my German research stay and regular summer schools were of indispensable value to my current and future work.

I am greatly thankful for the institutions that made my studies possible financially. TOSST provided continuous funding for four years to cover tuition, living expenses and TOSST related activities such as summer schools and

international research. The DAAD (Deutscher Akademischer Austausch Dienst) supported me for one year and the SKS (Stiftung für Kanadastudien) enabled the transition from Germany to Canada by providing funding for travel-related expenses such as a study permit and flight tickets.

A very special thanks goes to my family. Glen, Anni and Flyn Ratten have motivated me throughout my research and have continuously showed me their unquestionable love, support and understanding. I am very thankful that they intentionally or unintentionally steered my thoughts away from work to the other important parts in their lives and my life when I needed it.

My final thank you goes to my parents, Gudrun and Harm-Peer Zimmermann, for continuously enabling my studies. They have always provided me with unconditional love as well as emotional and financial support through which my plans have become reality.

CHAPTER 1: INTRODUCTION

Marine microbial communities play significant roles in the global cycling of nutrients, including the nitrogen cycle. The most common form of nitrogen on earth is N_2 gas, which makes up ca. 78% of the atmospheric gases and is dissolved throughout marine environments. Nevertheless, N_2 gas is inaccessible as a source of nitrogen for the vast majority of living organisms, who require forms of fixed nitrogen for their growth. During N_2 fixation, one molecule of N_2 gas is converted into two ammonia molecules by nitrogenase (equation 1), which requires 16 ATPs (Burgess and Lowe 1996):

$$N_2 + 8\ H^+ + 8\ e^- \rightarrow 2\ NH_3 + H_2 \quad (1)$$

Biological N_2 fixation is an important source of bio-available nitrogen in the ocean, where primary production, particularly in the oligotrophic subtropical gyres, is limited by nitrogen availability (Moore et al. 2013). N_2 fixation also plays an important role in replenishing fixed nitrogen lost through the pathways of denitrification and anammox (Gruber 2008; Codispoti 2007; Hamersley et al. 2007; Codispoti et al. 2001; Ingall et al. 1994). Our current knowledge of nitrogen balance still lacks major components as seen for example in the calculations of global oceanic inputs and outputs of fixed nitrogen species. There are discrepancies between stable isotope measurements in ocean sediments and the estimation of N_2 fixation rate measurements; the former suggests a balanced

budget over last 3000 years (Altabet 2007), whereas the later indicates an overall loss of fixed nitrogen from the oceans over the same time period (Codispoti 2007; Mahaffey et al. 2005). Two key reasons have been identified as to why calculations of biological N_2 fixation may underestimate the total input: Firstly, the diazotrophic community, a specific guild of bacteria and archaea that can perform biological N_2 fixation, has not been investigated in its entirety. Particularly the understanding of the non-cyanobacterial diazotrophs is lacking. Secondly, there has been a methodological error in N_2 fixation rate measurements that has led to an underestimation of N_2 fixation rates by up to 62% (Grosskopf et al. 2012). This leaves a large gap in our knowledge of the marine diazotrophic community, highlighting the need for more research into the diversity and metabolism of these organisms.

Nitrogenase

There are three types of nitrogenases, which are distinguishable by their metal cofactor requirements. The Mo-Fe nitrogenase is the most widely distributed type of nitrogenase among diazotrophs and therefore also the most commonly investigated. The other two nitrogenases contain either iron only (Fe-Fe, *anf* genes) or vanadium (V-Fe, *vnf* genes). They are less efficient than the Mo-Fe nitrogenase and are thought to be of importance in regions of Mo limitation (McRose et al. 2017).

The Mo-Fe nitrogenase is made up of two subunits: the homodimeric Fe protein (dinitrogenase reductase) that transfers electrons from the reducing agent to the second subunit, the heterodimeric Mo-Fe protein, which then reduces N_2 gas to NH_3 (Figure 1.1). The two subunits are coded for by the *nifHDK* genes. A ca. 360 bp fragment of the *nifH* gene (coding for the Fe protein) has become the established standard for exploring the diversity of diazotrophs on a molecular level (Zehr et al. 2001).

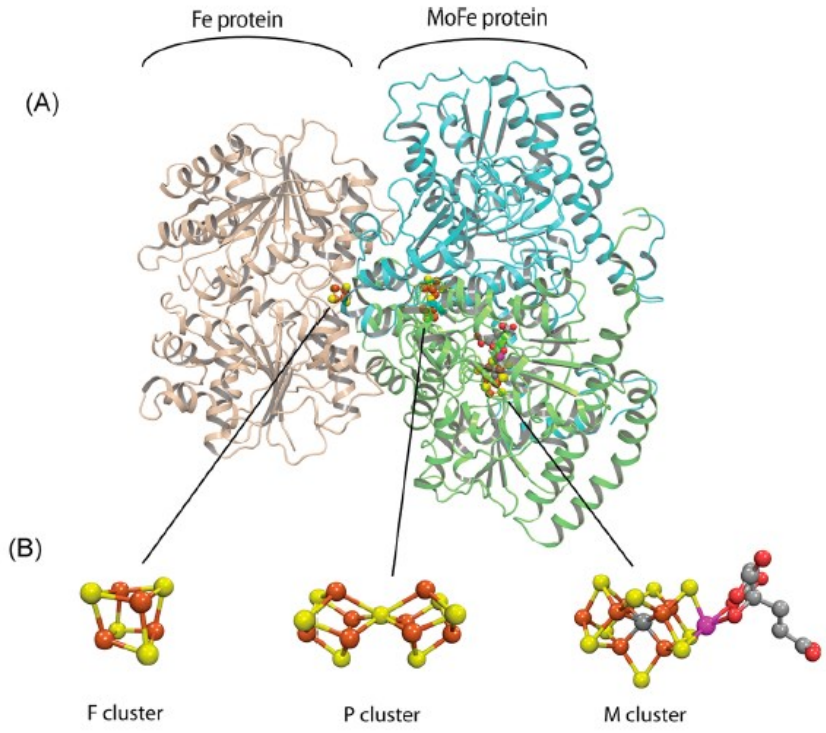


Figure 1.1: Molybdenum nitrogenase (from Hoffmann et al. 2014).

(A) One catalytic half of the Fe protein:Mo-Fe protein complex with the Fe protein homodimer shown in tan, the Mo-Fe protein α subunit in green, and the β subunit in cyan. (B) Space filling for the 4Fe−4S cluster (F), P-cluster (P), and Fe-Mo-co (M).

Initial investigations into the diversity of the *nifH* gene identified four major clusters. Cluster I was divided into two sub-clusters either dominated by cyanobacteria (sub-cluster Ic) or proteobacteria and some V-Fe (sub-cluster Ip). Cluster II contains archaeal and Fe-Fe nitrogenases and cluster III includes mainly anaerobic nitrogenases. Cluster IV is a group of enzymes of different function, but *nifH*-like sequences (Zehr et al. 2003).

Most studies on marine N_2 fixation at the organismal and molecular level have been focussed on cyanobacteria because of the predominance of *Trichodesmium*, unicellular cyanobacteria (*Candidatus* Atelocyanobacterium thalassa, *Cyanothece* and *Crocosphaera*) and symbiotic *Richelia* in the tropical surface oceans, while research on other diazotrophs have lagged behind (Luo et al. 2012). With the recent emergence of high-throughput sequencing and the expansion of study sites throughout all marine environments, there is increasing evidence supporting the importance of non-cyanobacterial diazotrophs in the ocean (Bombar et al. 2016; Langlois et al. 2015; Farnelid et al. 2011). The most commonly detected *nifH* sequences outside the tropical surface ocean, including in temperate, deep and O_2-depleted water masses, are of proteobacterial origin (Fernandez et al. 2015; Loescher et al. 2014; Farnelid et al. 2011; Hewson et al. 2007; Bird et al. 2005). These findings dispute the initial doubts concerning the presence of non-cyanobacterial diazotrophs in the ocean, which assumed that non-cyanobacterial *nifH* sequences stemmed primarily from contamination of PCR reagents (Izquierdo et al. 2006; Goto et al. 2005; Zehr et al. 2003b). However, studies outside the tropical surface ocean, especially at higher

latitudes, are sparse. Additionally, the initial methodology for high-throughput *nifH* sequencing limited the read length to ca 150 bp, but this has since improved to include a 360 bp segment (Cheung et al. 2016; Bentzon-Tilia et al. 2015; Xiao et al. 2015; Farnelid et al. 2013; Farnelid et al. 2011). Hence, there is increasing evidence suggesting that non-cyanobacterial diazotrophs play a role within the oceanic microbial communities. Additional research is needed to investigate the role and importance of non-cyanobacterial diazotrophs in the marine nitrogen cycle.

Metabolism

Based on the susceptibility of the nitrogenase to oxidation damage as well as phylogenetic analysis of *nif* genes, it has been proposed that the enzyme evolved before the oxygenation of the atmosphere (ca 2.4 – 2.2 Ga ago; Summons et al. 1999; Falkowski 1997). Diazotrophs have evolved mechanisms to reduce O_2 concentrations in the proximity of the nitrogenase during N_2 fixation periods to avoid oxidation of the reaction centre. These O_2 evasion mechanisms have been extensively studied in cyanobacteria, because the O_2 evolved during photosynthesis could directly oxidize the nitrogenase (Zehr 2011). The adaptation mechanisms include cell differentiation into O_2-excluding heterocysts (*Richelia*; Haselkorn 2007; Jahson et al. 1995), the loss of the genes coding for the O_2-evolving complex of the photosynthetic apparatus (eukaryote- symbiont *Candidatus* A. thalassa; Zehr et al. 2008) and the separation of N_2 fixation and photosynthesis over a diurnal cycle in unicellular cyanobacteria (*Cyanothece* and

Crocosphaera; Bandyopadhyay et al. 2011; Mohr et al. 2010; Reddy et al. 1993). However, unlike cyanobacteria, there are few established cultures representing non-cyanobacterial diazotrophs. Therefore, it is difficult to assess the adaptive mechanisms that have evolved to counteract the detrimental effects of O_2 in those organisms. From observations in terrestrial microorganisms, several mechanisms have been proposed: conformational changes and protein-complex protection of the nitrogenase (Schlesier et al. 2015; Moshiri et al. 1995), respiratory protection (Inomura et al. 2017; Paulus et al. 2012; Poole and Hill 1997), synthesis of reducing equivalents (Thorneley and Ashby 1989) and bacterial interactions such as cell-to-cell clumping and flocculation (Bentzon-Tilia et al. 2015; Bible et al. 2015; Dingler et al. 1988; Dingler and Oelze 1987). The association of N_2 fixation with micro-anaerobic environments on organic aggregates, where intense respiration lowers O_2 concentrations, has also been proposed as an O_2-evading mechanism in marine non-cyanobacterial diazotrophs (Riemann et al. 2010). Currently, we do not understand the O_2-evasion mechanisms of non-cyanobacterial diazotrophs, nor do we know the environmental conditions in which they are actively fixing N_2 and the magnitude of their N_2 fixation rates. Cultured representatives of non-cyanobacterial diazotrophs that are abundant in the marine environment are needed to determine their potential contribution to oceanic N_2 fixation. Studies of the few non-cyanobacterial diazotrophs already isolated have shown that their metabolisms are very diverse and extrapolation from only a few microorganisms would likely result in erroneous conclusions (Bentzon-Tilia et al. 2015; Fernendez et al. 2015; Riemann et al. 2010).

Outlook

Studies of diazotrophs and entire marine microbial communities are of importance in the context of the changing oceans. Microbial communities respond rapidly to changes in their environment and they are most likely already adapting to the changes in the oceans brought about by climate change (Doney et al. 2012; Giovannoni and Vergin 2012; Wright et al. 2012; Diaz and Rosenberg 2008; Arrigo 2005; Price and Sowers 2004; Pomeroy and Wiebe 2001). The biological activity of marine microbial communities is responsible for a large proportion of the global cycling of nutrients, and it is vital to understand their integrated response to climate change if we hope to predict future feedback mechanisms (Zehr and Kudela 2011; Falkowski et al. 2008; Kirchman 2000).

Diazotrophs play an important role in the cycling of nutrients through their metabolism, but also through the fertilizing action N_2 fixation takes in the ocean. This thesis addresses some of the current knowledge gaps related to diazotrophic distribution, diversity and metabolism using cultivation and culture-independent methods to improve our understanding of marine diazotrophs and the implications climate change might have on their distribution and diversity. The following research questions were addressed:

- What environmental parameters drive diazotrophic distribution in the tropical North Atlantic Ocean (Chapter 2)?
- What is the structure of diazotrophic communities and how does it vary throughout the Atlantic Ocean (Chapter 3)?
- What metabolic roles do diazotrophs play within the microbial community (Chapter 4)?
- What is the distribution, abundance and potential metabolism of a newly isolated heterotrophic diazotroph (Chapter 5)?
- How does the marine microbial community, including the diazotrophic community, in the Bedford Basin change throughout the year (Chapter 6)?

CHAPTER 2: SOURCES OF IRON AND PHOSPHATE AFFECT THE DISTRIBUTION OF DIAZOTROPHS IN THE NORTH ATLANTIC

Published Manuscript:

Ratten, J.M., LaRoche, J., Desai, D.K., Shelley, R.U., Landing, W.M., Boyle, E., Cutter, G.A. and Langlois, R.J., 2015. Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. Deep Sea Research Part II: Topical Studies in Oceanography, 116, pp.332-341.

Contribution of authors:

Jenni-Marie Ratten:	Sample extraction and processing, data analysis, drafting of manuscript
Julie LaRoche:	Planning and discussion of the manuscript
Dhwani K. Desai:	Support with statistical analysis
Rachel U. Shelley:	Provided aerosol data
William M. Landing:	Provided aerosol data
Ed Boyle:	Collected the samples
Gregory A. Cutter:	Provided nutrient data
Rebecca J. Langlois:	Sample extraction, planning and discussion

2.0. Abstract

Biological nitrogen fixation (BNF) supplies nutrient depleted oceanic surface waters with new biologically available fixed nitrogen. Diazotrophs are the only organisms that can fix dinitrogen, but the factors controlling their distribution patterns in the ocean are not well understood. In this study, the relative abundances of eight diazotrophic phylotypes in the subtropical North Atlantic Ocean were determined by quantitative PCR (qPCR) of the *nifH* gene using TaqMan probes. A total of 152 samples were collected at 27 stations during two GEOTRACES cruises; Lisbon, Portugal to Mindelo, Cape Verde Islands (USGT10) and Woods Hole, MA, USA via the Bermuda Time Series (BATS) to Praia, Cape Verde Islands (USGT11). Seven of the eight diazotrophic phylotypes tested were detected. These included free-living and symbiotic cyanobacteria (unicellular groups (UCYN) A, B and C, *Trichodesmium*, the diatom-associated cyanobacteria *Rhizoselinia-Richelia* and *Hemiaulus-Richelia*) and a γ-proteobacterium (Gamma A, AY896371). The *nifH* gene abundances were analyzed in the context of a large set of hydrographic parameters, macronutrient and trace metal concentrations measured in parallel with DNA samples using the PRIMER-E software. The environmental variables that most influenced the abundances and distribution of the diazotrophic phylotypes were determined. We observed a geographic segregation of diazotrophic phylotypes between east and west, with UCYN A, UCYN B and UCYN C and the *Rhizosolenia-Richelia* symbiont associated with the eastern North Atlantic (east of 40°W), and *Trichodesmium* and Gamma A detected across the basin. *Hemiaulus-Richelia*

symbionts were primarily found in temperate waters near the North American coast. The highest diazotrophic phylotype abundance and diversity were associated with temperatures greater than 22°C in the surface mixed layer, a high supply of iron from North African aeolian mineral dust deposition and from remineralized nutrients upwelled at the edge of the oxygen minimum zone off the north western coast of Africa.

2.1. Introduction

Biological nitrogen fixation (BNF) is an important source of biologically available nitrogen in the marine environment, as fixed forms of nitrogen are scarce in most open ocean surface waters. BNF is carried out by specific groups of Bacteria and Archaea called diazotrophs. In areas such as the oligotrophic subtropical gyres they provide the most significant source of fixed nitrogen (Duce et al. 2008; Karl et al. 2002; Vitousek and Howarth 1991). Over geological time scales, the magnitude of the global oceanic fixed nitrogen inventory has been determined by the balance between BNF and the combined nitrogen loss processes of denitrification and anaerobic ammonia oxidation (anammox; Altabet 2006; Codispoti 2006).

Until a decade ago, it was believed that most of the BNF in the ocean was performed by the large, surface bloom-forming *Trichodesmium* , a non-heterocystous, filamentous cyanobacterium, and by symbiotic associations between diatoms and the diazotroph *Richelia* sp. (Foster et al. 2007). Phylogenetic studies using *nifH*, the gene encoding for the iron protein sub-unit of

the nitrogenase enzyme have revealed a much more diverse diazotrophic flora that includes unicellular and symbiotic cyanobacteria, heterotrophic bacteria and archaea, all potentially contributing significantly to global oceanic BNF (Turk et al. 2011; Langlois et al. 2005; Zehr et al. 1998). High-throughput next generation sequencing studies have further enriched our knowledge of diazotroph phylogenetic diversity, and have identified the presence of unexplored groups of heterotrophic diazotrophs throughout the world's oceans (Farnelid et al. 2011).

Although the abundance of the diazotrophs *Trichodesmium* and *Richelia* can be determined by microscopy counts, many other diazotrophic unicellular cyanobacteria and heterotrophic bacteria in marine microbial communities cannot be visually identified with certainty by microscopy alone. Microscopic images of the elusive UCYN A, one of the most widely distributed diazotrophic cyanobacteria, have been obtained only recently (Krupke et al. 2013; Thompson and Zehr 2013). To date, most oceanic heterotrophic diazotrophs are known only by their *nifH* sequences. To further complicate the matter, the abundance of diazotrophs is generally several orders of magnitude lower than the dominant phytoplankton and bacterioplankton (e.g. *Prochlorococcus* and *Pelagibacter*). This presents a challenge for detection and cultivation techniques. Quantitative PCR (qPCR) and TaqMan probes have been used to circumvent some of these difficulties (Langlois et al. 2008), allowing the quantitative detection of diverse phylogenetic clades defined by specific *nifH* sequences. This approach has already yielded valuable information on *nifH* phylotype distributions and

abundances in the Pacific (Moisander et al. 2010; Church et al. 2008; Goebel et al. 2007) and Atlantic Oceans (Turk et al. 2011; Langlois et al. 2008).

Diazotroph distribution has been utilized to estimate areas of BNF and model the factors controlling BNF. However, the oceans remain vastly under sampled with respect to diazotroph abundance, distribution and community structure (Fernández et al. 2013; Luo et al. 2012), making it problematic to validate model-based predictions concerning the fate of BNF in a changing ocean (Sohm et al. 2011; Monteiro et al. 2010; Goebel et al. 2007). It is therefore important to collect additional data on the diazotroph distribution in regions that are currently undersampled in order to better constrain the factors controlling BNF.

Environmental parameters such as temperature, availability of phosphate, water column stability, upward diffusive fluxes of nutrients, light and input of iron (Fe) via atmospheric mineral dust deposition have all been proposed as factors controlling the distribution of diazotrophs (Fernández et al. 2013). Although detected in almost every oceanic environment, diazotrophs are most abundant in the warm tropical and subtropical oceans where fixed nitrogen is depleted in surface waters (Moisander et al. 2010; Stal 2009; Church et al. 2008; Langlois et al. 2008). In contrast to primary producers, diazotrophs are not limited by fixed nitrogen availability: instead both phosphorus and dissolved Fe availability have been implicated in the control of the geographical distribution of diazotrophs and BNF (Moore et al. 2009; Mills et al. 2004; Karl et al. 2002; Falkowski, 1997). In the oligotrophic subtropical North Atlantic gyre, mineral dust deposition is the most significant source of dissolved Fe to the surface of the ocean (Conway and

John 2014; Jickells et al. 2005; Gao et al. 2001). In the eastern tropical Atlantic, between the Cape Verde Islands and the north west African coast, upwelled regenerated nutrients from the sub-surface oxygen minimum zone (OMZ) are an additional potential source of macro-(N, P, Si) and micro-nutrients (e.g. Fe, Co) to the surface layers (Fitzsimmons et al. 2013; Rijkenberg et al. 2012; Noble et al. 2012; Bergquist and Boyle 2006a).

We used qPCR and eight phylotype-specific TaqMan probes and primer sets, representing the most commonly occurring marine diazotrophs in the surface Atlantic Ocean (Langlois et al. 2008) to estimate *nifH* abundances in an East–West transect across the subtropical North Atlantic Ocean. We compared the distribution and relative abundance of *nifH* phylotypes with hydrographic parameters, macronutrients and trace metal distributions from the surface to 400 m, as well as aerosol aluminium (Al) and Fe concentrations. This was possible through coordinated sampling of nucleic acids and a suite of trace metals dissolved in the water column and aerosols during the 2010 and 2011 US GEOTRACES research cruises.

2.2. Materials and Methods

2.2.1. Cruise track and sample collection

Samples for measuring *nifH* gene abundances for qPCR were collected during two GEOTRACES cruises (USGT10 and USGT11) that took place in the subtropical North Atlantic Ocean from October 16th to November 2nd 2010 and

from November 7th to December 10th 2011, respectively (Figure 2.1). The cruise track (Figure 2.1) included stations at the Bermuda Atlantic Time-series (BATS) site, Cape Verde Ocean Observatory (CVOO) site and the mid-Atlantic ridge (MAR). Seawater samples for *nifH* qPCR were collected from the conventional CTD/rosette at six depths per station ranging from 2 to 1000 m. Immediately after collection 1 – 2 L of seawater were vacuum filtered onto 0.22 µm Durapore filters (Millipore) to collect the natural microbial communities. The filters were stored at -80°C until analysis in the laboratory. In total, 152 samples were collected from 27 stations with an average of 6 depths per station. Up to three samples were collected in the surface mixed layer (SML) at all of the stations sampled. A broad suite of trace metals and other macronutrients were sampled during these two US GEOTRACES cruises (Deep-Sea Research II special issue), enabling the analysis of the nucleic acid-derived *nifH* abundance measurements within the context of a large data base of chemical and hydrographic parameters.

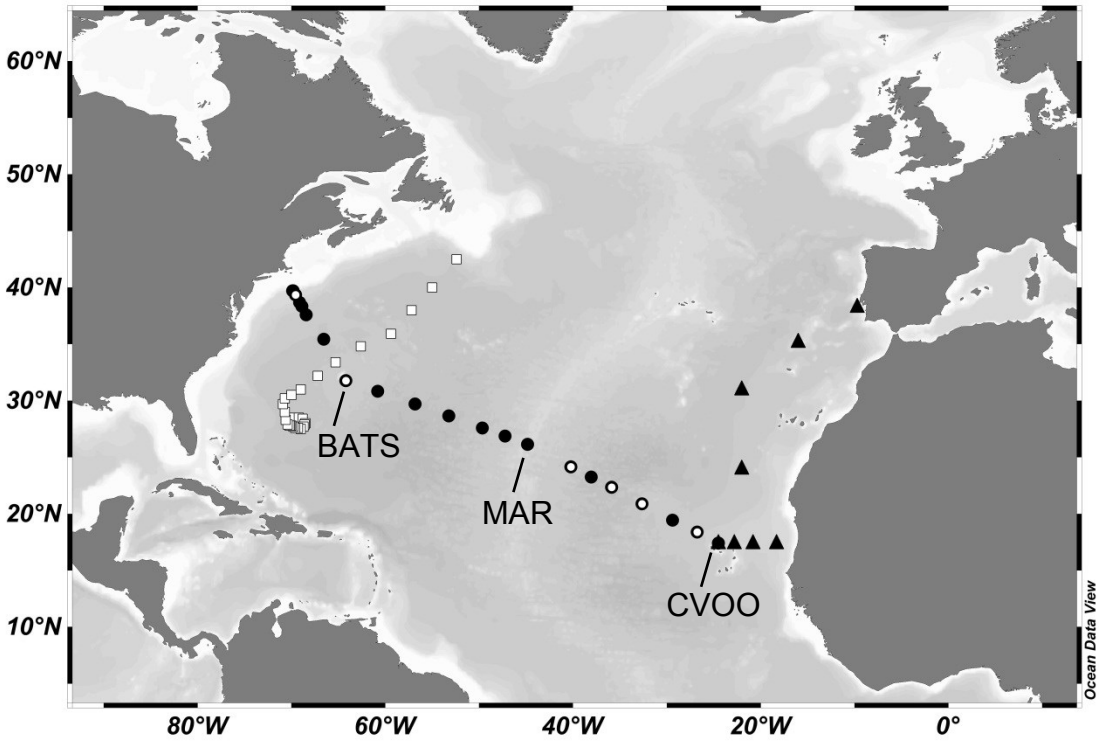


Figure 2.1: Cruise tracks and stations of USGT10 (triangles) and USGT11 (circles) in 2010 and 2011.

Labelled are the time series stations BATS and CVOO as well as the mid-Atlantic ridge (MAR). Stations with very high *Trichodesmium nifH* abundances are indicated with open circles. The track of tropical storm Sean (at BATS on 11 Nov. 2011) is overlaid in open squares.

2.2.2. DNA extraction and qPCR

In the laboratory liquid nitrogen-frozen filters were crushed with plastic homogenizers and incubated for 5 min with a 5 mg mL^{-1} lysozyme in TE buffer solution. DNA was extracted using the All Prep RNA/DNA MiniKit (Qiagen) following the manufacturer's protocol, except that DNA was eluted twice with 40 µL TE buffer and incubated for 5 min before centrifuging. DNA was stored in small aliquots to avoid freeze/thaw cycles. DNA concentrations were determined using the Quant-iT PicoGreen dsDNA reagent (Molecular Probes, Life Sciences).

The abundances of eight *nifH* phylotypes were determined by qPCR using the specific TaqMan probes and primers for Het1 (*Rhizosolenia-Richelia* symbionts; Church et al. 2005) and Het2 (*Hemiaulus-Richelia* symbiont; Foster et al. 2007), *Trichodesmium* , UCYN A (*Candidatus* Atelocyanobacterium thalassa), UCYN B (*Crocosphaera*), UCYN C (*Cyanothece*), Gamma A (gamma-proteobacteria A) and cluster III (Langlois et al. 2008). Universal TaqMan master mix and concentrations of primers, probes and BSA were mixed as in Langlois et al. (2008) in a reaction volume of 25 mL, which included either 5 µL of plasmid standard, DNA sample or PCR water as template. Plasmid standards, samples and no-template controls were run in duplicate on the Roche Light Cycler 480 using clear 384-wellplates. Samples were amplified using the following program: 95°C for 10 min, 45 cycles of [95°C for 15 s, 60°C for 1 min]. Data was collected at 60°C. A ramp of 1.6°C s^{-1} was used at each step.

Amplification curves were analyzed using LinReg (version 2013.0; Ramakers et al. 2003). Average primer efficiencies (Langlois et al. 2012) were 97% for *Rhizosolenia-Richelia* symbiont, *Hemiaulus-Richelia* symbiont and UCYN A, 91% for *Trichodesmium*, UCYN B and UCYN C, 92% for Gamma A and 95% for cluster III. As it is not yet known how the *nifH* copy numbers relate to diazotroph biomass or cell density, the qPCR results are reported throughout the manuscript as *nifH* copies mL^{-1} and represent the number of *nifH* copies detected in environmental DNA samples in a known volume of seawater. All phylotypes except cluster III were detected. Hence cluster III is not included in the analysis.

2.2.3. PRIMER-E analysis

2.2.3.1. Preparing the matrices

The *nifH* gene abundances of the seven detected diazotrophic phylotypes (*Rhizosolenia-Richelia* symbiont, *Hemiaulus-Richelia* symbiont, *Trichodesmium* , UCYN A, UCYN B, UCYN C and Gamma A) and their corresponding environmental variables were analyzed in all 152 samples or a subset of the surface mixed layer samples (SML) only (52 samples) using the PRIMER-E software V.6 (Clarke and Gorley 2006). The SML was defined as waters from the surface down to 1 m above the mixed layer depth, calculated from temperature and potential density profiles (Mariko Hatta, personal communication). Environmental metadata was obtained from BCO-DMO (Biological and Chemical Oceanography Data Management Office; links to all data sets are listed in Supplemental Table 1). The dataset was first divided into a community data matrix (containing gene abundances of the *nifH* phylotypes) and a corresponding matrix of environmental measurements (including dissolved metal concentrations, nutrients, organic material and physical parameters).

Missing values in one or more environmental variables resulted in the deletion of the entire sample from the database and hence, such samples were not included in the subsequent statistical analysis. A correlation matrix (draftsman's plot) was generated for each pair of environmental variables. Only one variable was retained from pairs with a correlation > 75. This resulted in a final dataset comprised of 64 samples divided into a subset of samples collected within the

SML (37 samples) and a subset of samples originating from below the SML (27 samples).

The environmental variables used for the principal component analysis were expressed on broadly different scales precluding a direct comparison without biasing of the results. In order to derive meaningful distances between samples using Euclidian distances, we first square-root transformed or log transformed the variables that covered several orders of magnitude to bring them within a common numerical range. This generated values all ranging within 4 orders of magnitude, allowing variables to be compared without biases. Each variable of the environmental matrix was then normalized by subtracting their mean and dividing by their standard deviation prior to further analysis. The *nifH* phylotype abundance data was log-transformed and compared using Bray-Curtis Similarities, a similarity (or distance) measure that ignores joint absences of variables between samples.

2.2.3.2. Multivariate analysis pipeline

For all phylotype matrices a BEST (Bio-Env+Stepwise) test was carried out. This test determines which environmental variables best explain the microbial community composition. The comparison was carried out between the transformed environmental matrix and the Bray–Curtis similarities of the phylotype data set. A combination of variables which showed the maximum correlation with the phylotype distribution was identified for further analysis (Supplemental Table 2) using LINKTREE (linkage tree), a program that describes

the best way to split the samples into groups based on a threshold value for each environmental variable (for example group1 > 0.4 μM PO_4^{3-} > group 2).

Principle component analysis (PCA) was performed with the environmental matrix of the SML samples. The first three components of the PCA captured 85.2% (PC1 = 23.4%, PC2 = 51.3%, PC3 = 10.5%) of the variance. Based on the clustering obtained in the PCA plots and on the results of the BEST/LINKTREE test and utilizing information about known drivers of diversity, the samples were categorized into groups. These categories included geographical location (east or west of 40°W, north or south of 30°N), nutrient concentration (high, low), trace metal concentrations in the water column and in aerosols (high, low), dust origin (European, North African/Saharan, Marine, North American) and rain (present, absent). For variables with continuous data, high and low concentrations were defined by a threshold derived from the published peer-reviewed literature. If no definition could be found in the literature, the variables were categorized based on an evaluation of the present dataset (LINKTREE analysis) and literature values (Table 2.1).

An ANOSIM (analysis of similarity) test was utilized to compare the diazotrophic communities within these predefined groups and to determine whether the distribution of the *nifH* phylotypes were significantly different between the predefined groups for each environmental variable (Table 2.1). The Bray-Curtis similarities of the log-transformed diazotroph community data (based on *nifH* counts of the various phylotypes at each site) was used for the ANOSIM.

For each categorization that showed a positive ANOSIM test ($p < 0.05$), the discriminating phylotypes in the groups of this categorization (e.g. high and low aerosol concentration) were identified using the SIMPER (similarity percentages) routine (Table 2.2; the three most influential phylotypes are highlighted in bold and italics).

2.3. Results

2.3.1. Distribution of seven diazotrophic phylotypes during the US Atlantic GEOTRACES cruises

Although *nifH* phylotypes were detected throughout the water column, they were most abundant in the surface mixed layer (SML; Figure 2.2). Along the east–west transect, the average sum of all *nifH* phylotypes (*nifH* copies mL^{-1}) in the SML was highest close to the Cape Verde Ocean Observatory (CVOO) on the eastern side of the transect(> 100 *nifH* copies mL^{-1}), dropping off around the MAR and rising again on the western side of the basin to > 100 *nifH* copies mL^{-1} (Figure 2.3A). The mean and standard error SML values for the nutrients and for *nifH* abundances are plotted in Figure 2.3 and Figure 2.4. The large standard errors of the mean *nifH* abundances for some phylotypes are a result of depth profiles within the SML that exhibit surface (e.g. *Trichodesmium*) or subsurface (e.g. diatom-associated cyanobacteria) maxima. The structure in the depth profiles of the *nifH* phylotypes can also be seen in Supplemental Figure 1 and Supplemental Figure 2 that shows examples of vertical depth profiles for station with high variability within the SML. The Shannon diversity index (a measure of abundance

and evenness of species present) showed a similar trend, with high diversity of *nifH* phylotypes in the eastern basin and lower diversity observed in the center of the gyre (Figure 2.3B). Although the diversity and abundance of *nifH* phylotypes varied across the Atlantic basin, temperature and N* (N* = NO_3^- - $16PO_4^{3-}$ + 2.9 µM; Gruber and Sarmiento, 1997) remained relatively constant within the gyre, at 25.2°C and 2.8 µM, dropping at the western continental shelf edge to19°C and 1.9 µM, respectively (Figure 2.3C and Figure 2.3D). N* calculates the deviation of the nitrate:phosphate ratio from the Redfield Ratio of 16:1 (Redfield 1934) and has been widely used as a quasi-conservative tracer of nutrient remineralization processes in the ocean. In our study, N* was relatively high and positive throughout the transect, consistent with a gain in fixed N through BNF. The decrease in N* observed near the Cape Verde Islands and at the North American Coast reflected the peaks in phosphate at these two locations. Major nutrient concentrations (N, P, Si) concentrations were low in the SML and relatively constant throughout the transect; except for a tendency towards depletion of nitrate relative to phosphate in the western Atlantic and at CVOO. Phosphate concentrations were low across the gyre, averaging 7.5 nM. However, phosphate concentrations rose to16 – 74 nM near the American coast and to 21 nM east of CVOO (Figure 2.3E).

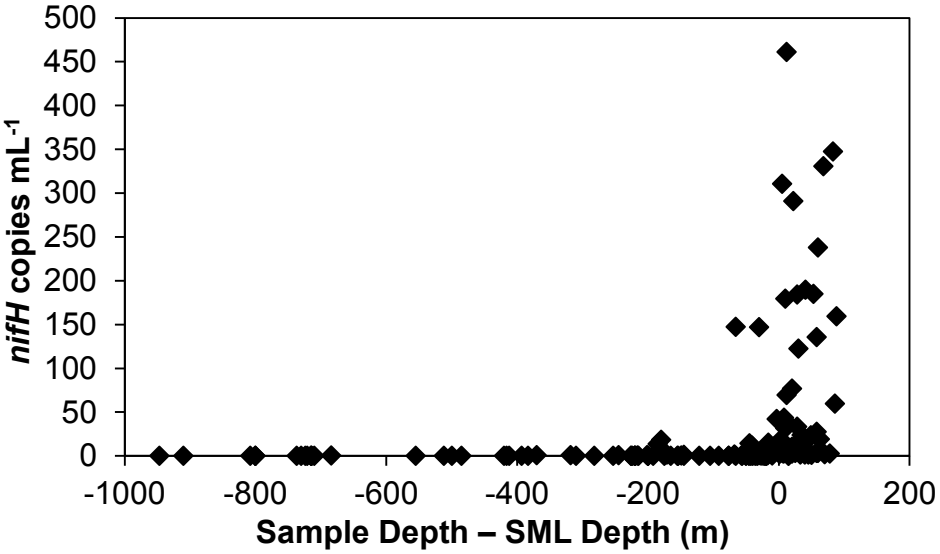


Figure 2.2: Abundances of total *nifH* copies mL^{-1} measured in all samples in relation to the difference between SML depth and sample depth.

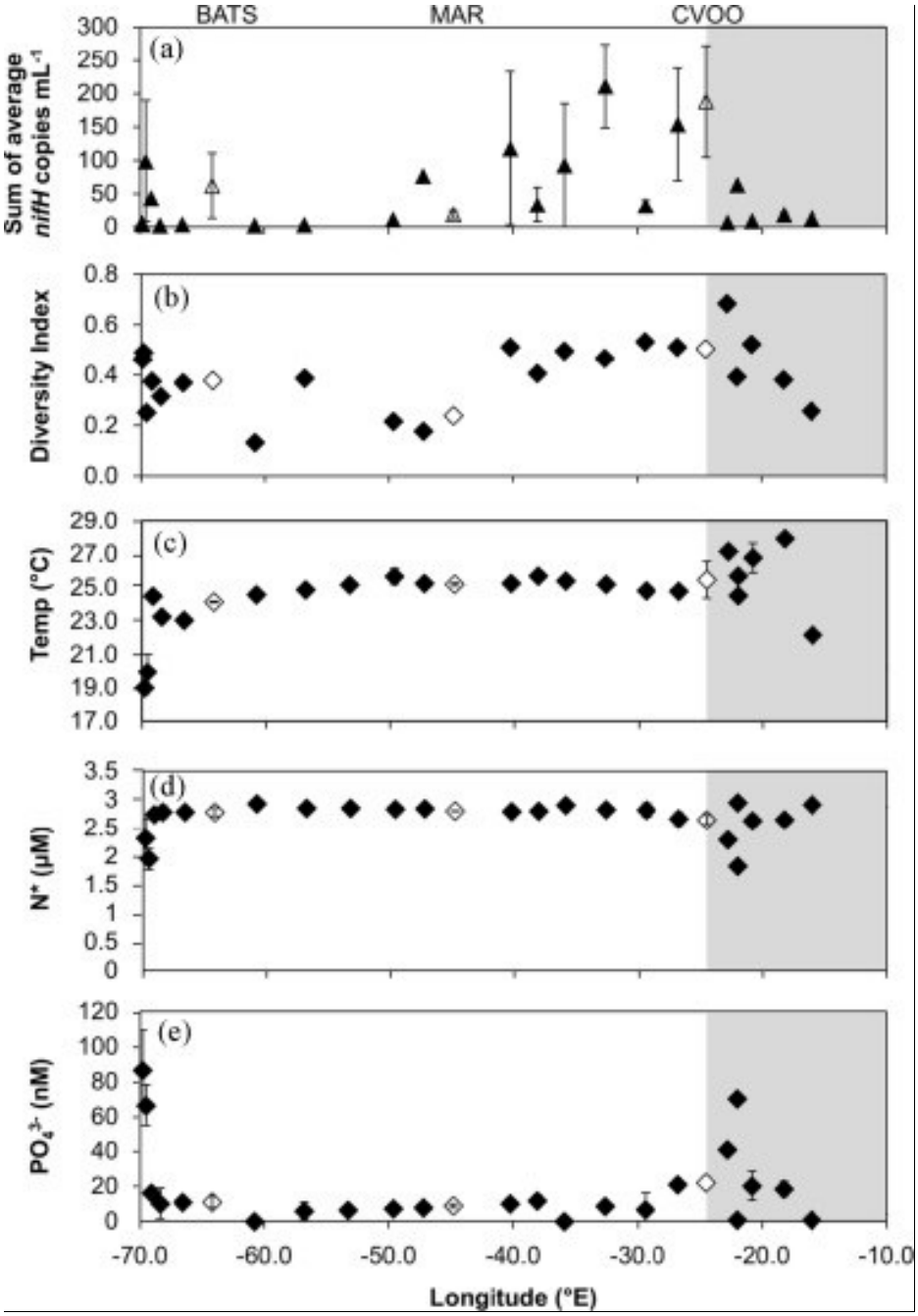


Figure 2.3: (A) Average sum of *nifH* copy numbers in the SML, (B) Shannon diversity index, (C) average temperature, (D) average N* and (E) average phosphate in the SML during cruises USGT10 (gray shaded) and USGT11 (no shading). Stations BATS, MAR and CVOO are indicated by open diamonds. If more than one sample was measured in the SML, averages are shown and the standard error is plotted.

Figure 2.4 shows the *nifH* gene copy numbers (copies mL^{-1}) for the seven detected *nifH* phylotypes. The most commonly detected phylotypes were *Trichodesmium* and UCYN A, reaching maximum abundances of 391 and 105 *nifH* copies mL^{-1} at individual stations, respectively (Figure 2.4A and Figure 2.4B). *Trichodesmium* was the most frequently detected phylotype throughout the transect with abundances greater than 150 *nifH* copies mL^{-1} detected at six stations (Figure 2.1) and contributed the most to the overall abundance of the sum of the *nifH* genes (Figure 2.3A and Figure 2.4A). *Rhizosolenia-Richelia* symbiont and UCYN C were the least abundant phylotypes with maximum abundances of 4 and 6 copies mL^{-1}, respectively at CVOO and Station 23 (Supplemental Figure 2). Although far less abundant than *Trichodesmium*, *Rhizosolenia-Richelia* symbiont and Gamma A phylotype distributions paralleled the variation in the *Trichodesmium* phylotype distribution (Figure 2.4C and Figure 2.4D). The unicellular cyanobacterial phylotypes UCYN A, B, and C were most abundant on the eastern side of the transect in the region between 30°W and 25°W, directly west of CVOO (Figure 2.4C, Figure 2.4D and Supplemental Figure 2), reaching significant *nifH* abundances below the SML (SML depth at 62 – 85 m). In contrast, the *Hemiaulus-Richelia* symbiont phylotype was mainly found in colder waters (19 – 25°C) near the American coast (Figure 2.4C).

We investigated the correlation between mineral aerosol concentrations (and therefore implied deposition) and *nifH* abundance in surface waters (Figure 2.4E). The high aerosol concentrations of Al and Fe observed on the eastern side of the transect (both elements exceeding 1000 ng m^{-3} at most sample points with maxima of 7620 and 5760 ng m^{-3} respectively at CVOO; Shelley et al. 2014)

coincided with the highest abundance of *nifH* copies mL^{-1} (Figure 2.3A and Figure 2.4E). Even though it has been shown that the North African dust samples have low fractional Fe solubility compared to aerosols originating from North America, the very high amount of North African dust that is transported to the eastern North Atlantic implies that the flux of soluble aerosol Fe would be higher at CVOO (Shelley et al. 2014). Analysis of Fe isotope ratios has indeed recently confirmed that Saharan dust is the dominant source of Fe to the North Atlantic (Conway and John 2014). A rapid decrease in aerosol Al and Fe concentrations coincided with the decrease in diversity and *nifH* abundance at 40°W.

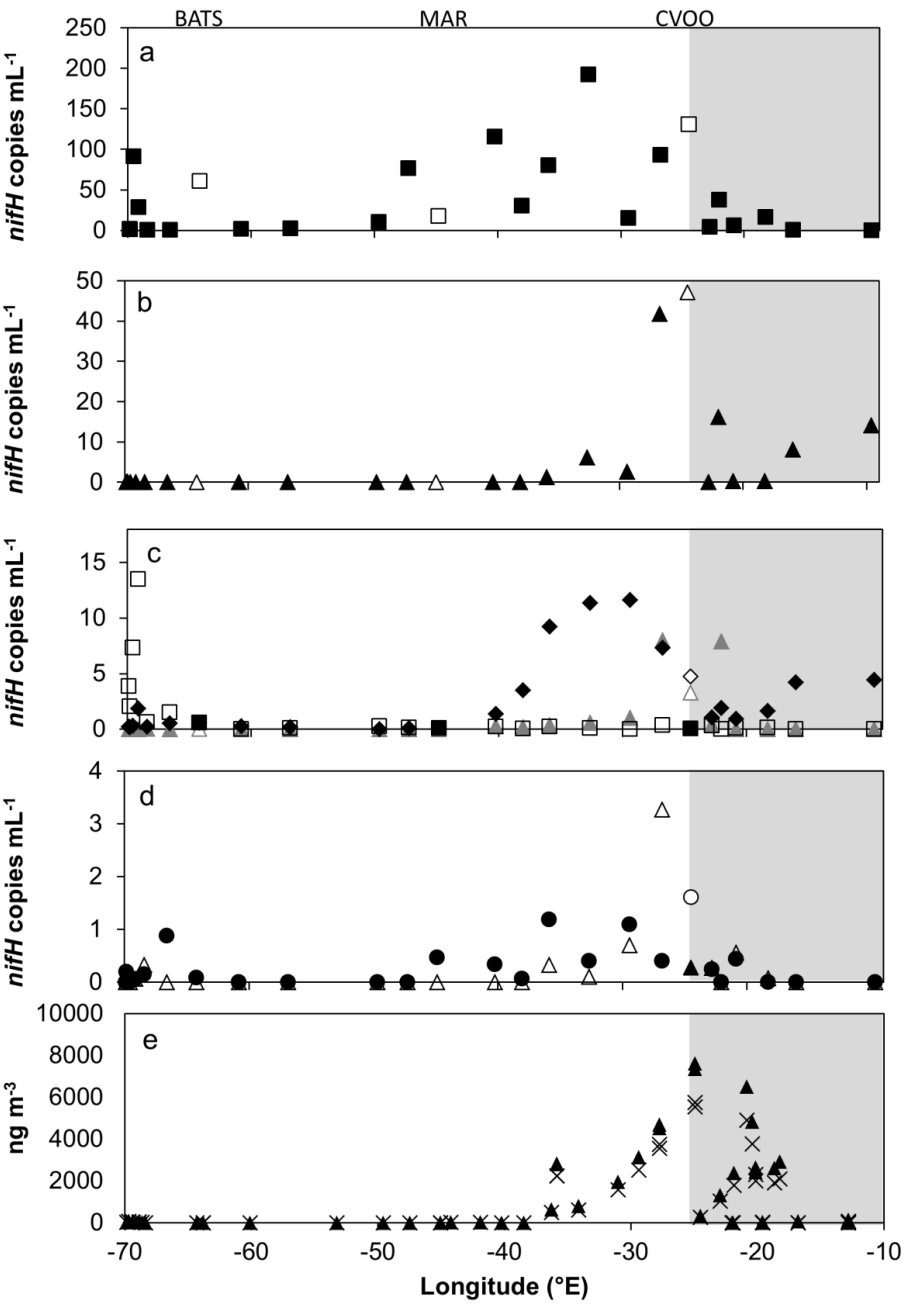


Figure 2.4: Copy numbers of *nifH* (mL^{-1}) of (A) *Trichodesmium*, (B) UCYN A, (C) Gamma A (black diamonds), UCYN B (grey triangles), *Hemiaulus-Richelia* (open squares) and (D) UCYN C (open triangles), *Rhizosolenia-Richelia* (black circles) in the SML. If more than one sample was taken from the SML, averages are shown and the standard error is plotted. (E) aluminum (black triangles) and iron (black crosses) concentrations from aerosol samples (ng m^{-3}) from USGT10 (gray shading) and USGT11 (no shading). Stations BATS, MAR and CVOO are indicated by white symbols.

2.3.2. Multivariate statistical analysis

The BEST test applied to the entire data set provided an initial identification of the environmental variables most relevant in determining the observed diazotrophic community composition. In total, 26 environmental variables were tested including dissolved inorganic nutrients and hydrographic parameters (Supplemental Table 1). The results of this analysis demonstrated the importance of certain characteristics of the SML as a determining factor for the community composition ($p = 0.01$; Supplemental Table 2). We therefore carried out the BEST analysis on two subsets composed of samples present above and below the SML referred to as SML or deep samples, respectively. In contrast to the SML samples, the deep sample subset showed no significant correlation, likely due to the very low *nifH* phylotype abundances measured in most deep samples ($p = 0.31$; Figure 2.2), and was not analyzed further.

For the SML, temperature, phosphate, and other environmental variables identified from the BEST analysis were used in a PCA (Figure 2.5). With the exception of Aeolian trace metals, none of the available dissolved trace metal data that were measured during the cruise showed significant influence on the distribution of the *nifH* phylotypes. This included the biologically relevant cofactors cobalt, vanadium, copper and zinc.

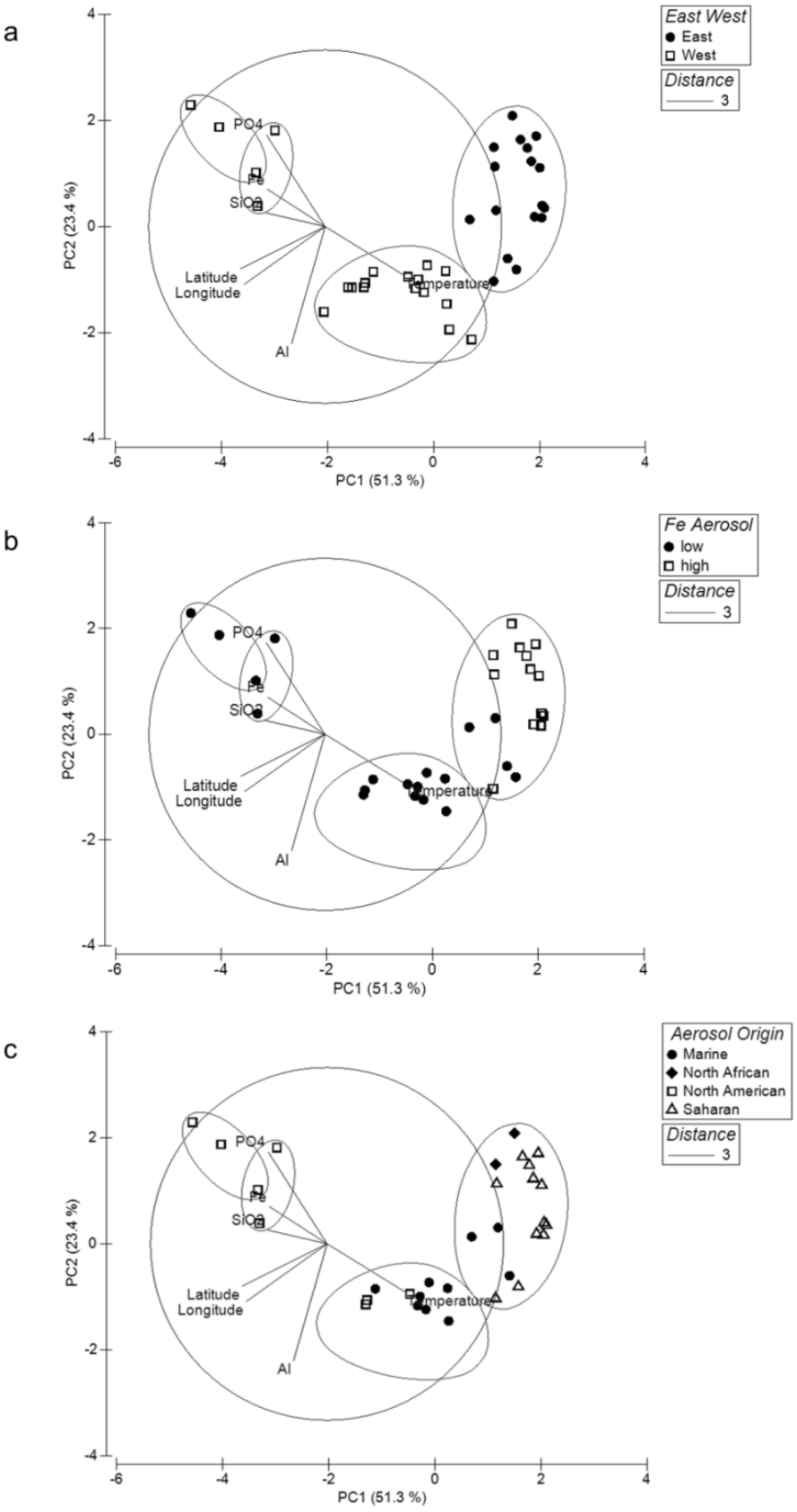

a
East West
East
West
Distance
3
PO4
Fe
SiO2
Latitude
Longitude
Al
Temperature
PC2 (23.4 %)
PC1 (51.3 %)
b
Fe Aerosol
low
high
Distance
3
PO4
Fe
SiO2
Latitude
Longitude
Al
Temperature
PC2 (23.4 %)
PC1 (51.3 %)
c
Aerosol Origin
Marine
North African
North American
Saharan
Distance
3
PO4
Fe
SiO2
Latitude
Longitude
Al
Temperature
PC2 (23.4 %)
PC1 (51.3 %)

Figure 2.5: Principal Components Analysis (PCA) of SML samples from USGT10 and USGT11 showing variables that contributed to significant clustering of samples.

Significant clusters are traced with a line representing a Euclidean-distance of 3 obtained from a hierarchical cluster analysis of the samples (significant clusters are labeled A–C). (A) Samples collected west of 40°W are indicated with open squares and east of 40°W with black circles;(B) and (C) show aerosol data provided by Shelley et al. (2014) with (B) high aerosol iron concentrations (above 50 ng m^{-3}) plotted as open squares, low iron aerosol concentrations (below 50 ng m^{-3}) plotted as black circles, and (C) aerosol origin as back trajectories over the past 5 days: Marine (black circles), North African (black triangles), North American (open squares).

The PCA conducted with the environmental variables identified with BEST resulted in three significantly different (ANOSIM, $p < 0.01$) clusters of samples (labeled A – C, Figure 2.5). The most important determining factors were east–west segregation, mineral dust concentration and nutrient upwelling of P and Fe. A and B were composed of samples collected east and west of 40°W, respectively, both characterized by high water temperatures and low macronutrient concentrations (Figure 2.5A). The eastern cluster (A) was dominated by high aerosol Fe concentrations of African origin. The large western cluster (B) was dominated by aerosols originating from marine and North American sources containing lower Fe concentrations (Figure 2.5B and Figure 2.5C).

Two overlapping western clusters (C) composed of samples collected near the North American coast, in waters with high nutrients and low temperature were not significantly different from each other, as determined by hierarchical clustering analysis (results not shown). The samples from the North American

coast were subjected to low aeolian Fe originating from North America rather than North Africa.

The original PCA plot was over laid with *nifH* phylotype abundances (Supplemental Figure 3). Some of the phylotypes were distributed evenly across the clusters (*Trichodesmium* and *Rhizosolenia-Richelia* symbionts), whereas the high abundances of UCYN A, UCYN B, UCYN C and Gamma A *nifH* copies coincided with the North African high aerosol cluster in the east of the basin (A); *Hemiaulus-Richelia nifH* copies were associated with the higher macronutrient and lower temperature conditions near the North American coast (Figure 2.5C; Supplemental Figure 3).

To further analyze whether these differential patterns of phylotype distribution had a statistical significance, we performed ANOSIM tests on the community matrix (i.e. the *nifH* phylotype abundances). The samples were divided into groups or categories as explained in the methods. Categorizations which showed a positive ANOSIM test ($p < 0.05$) were temperature, longitude and latitude, PO_4^{3-} (which correlated with NO_3^- and SiO_2), aerosol concentrations and aerosol origin (Table 2.1). Dissolved Al contributed to clustering in the PCA, but was not significant in the ANOSIM. Dissolved Fe concentrations were also not significant as seen previously in Table 2.1.

Table 2.1: Statistical comparison (ANOSIM) of SML *nifH* phylotype abundances with environmental variables.

	Threshold	R statistic[1)]	Significance level %
Physical parameters			
Temperature[2)]	22 °C	0.256	**0.5**
East – West	40°W	0.669	**0.1**
rain	presence	0.023	38.4
Nutrients (all low)			
N:16P ratio	1	0.723	5.9
Trace metals			
Dissolved Al[3)]	20 nM	0.053	27
Dissolved Fe[4)]	0.2 nM	-0.034	62.9
Dissolved Co	50 pM	-0.009	49.1
Dissolved Mn[5)]	2 nM	0.008	43.4
Dust			
Al aerosol[6)] (air sample)	50 ng m^{-3}	0.582	**0.1**
Fe aerosol[7)] (air sample)	50 ng m^{-3}	0.548	**0.1**
Aerosol origin		0.639	**0.1**

1) An R value with a significance level lower than 5% indicates that groupings are significantly different from each other. Significant variables are highlighted in bold.
2) Langlois et al. (2008)
3) Dammshäuser et al. (2011)
4) Moore et al. (2009)
5) Shiller (1997)
6) Buck et al. (2010)
7) Buck et al. (2010)

The phylotypes which contributed to significant differences between the groups were identified with the SIMPER routine and are listed in Table 2.2.

Trichodesmium and Gamma A, which were both distributed throughout the

transect, were not generally discriminating taxa, except in the case of the aerosol origin categories where higher *Trichodesmium* abundances were found in association with conditions where the air-mass back trajectories indicated that the aerosols did not have an obvious continental source (Shelley et al. 2014). UCYN A and UCYN B were in almost all groupings major contributors to differences and were significantly found east of the MAR in the area most influenced by high mineral aerosol concentrations near Cape Verde.

Table 2.2: Average log abundances of discriminatory SML *nifH* phylotypes that contributed to the overall dissimilarity between sample grouping pairs (dissimilarity/standard deviation > 1) determined using SIMPER. Group pairs are defined using the thresholds in Table 2.1. Species abundances that were significantly different in each group pair and contributed to the overall differences are in bold and italicized. Phylotypes with a dissimilarity/standard deviation below 1 are not displayed.

Sample groupings	*Rhizosolenia-Richelia*	*Hemiaulus-Richelia*	*Trichodesmium*	UCYN A	UCYN B	UCYN C	Gamma A
East of 40°W[1,2)]	***2.27***	1.33	4.22	***3.29***	***2.86***	1.76	3.73
West of 40°W	***0.90***	2.51	3.57	***0.32***	***0.39***	0.08	2.18
Marine aerosol	***0.96***	***1.76***	***3.64***	[3)]			
N. American aerosol	***1.05***	***3.31***	***3.30***				
Marine aerosol	0.96		***3.64***	***0.34***	***0.75***		2.11
N. African aerosol	1.18		***2.89***	***2.45***	***1.76***		3.02
N. American aerosol	1.05	***3.31***	3.30	***0.74***	***0.25***	0.00	
N. African aerosol	1.18	***1.39***	2.89	***2.54***	***1.75***	1.21	
Marine aerosol	***0.96***		3.64	***0.34***	***0.75***	0.00	2.11
Saharan aerosol	***2.64***		4.79	***3.50***	***3.04***	2.00	3.86
N. African aerosol	***1.18***	1.39	***2.89***	***2.54***	1.76	1.21	
Saharan aerosol	***2.64***	1.43	***4.79***	***3.50***	3.04	2.00	
N. American aerosol	1.05	***3.31***	3.30	***0.74***	***0.25***	0.00	
Saharan aerosol	2.64	***1.43***	4.79	***3.50***	***3.04***	2.00	
high Fe aerosol[4)]	2.47	1.48	4.50	***3.62***	***2.99***	***2.03***	3.75
low Fe aerosol[5)]	1.04	2.35	3.56	***0.49***	***0.56***	***0.00***	2.35
High Temperature[6)]	1.52	***1.79***	***3.91***	***1.59***	1.61		2.87
Low Temperature[7)]	1.03	***3.60***	***3.35***	***1.09***	0.00		2.36

1) Compared abundances are separated by dashed line.
2) The three major contributors to differences are printed in bold+italic
3) Not contributing phylotypes and phylotypes with a Dissimilarity/Standard Deviation below 1 are not displayed
4) Above 50 ng m^{-3}
5) Below 50 ng m^{-3}
6) Above 22 °C
7) Below 22 °C

2.4. Discussion

2.4.1. The diazotroph community structure along an east–west transect in the North Atlantic Ocean

It is now well established that the marine diazotrophic community is much more diverse than previously thought (Farnelid et al. 2011; Zehr et al. 1998). However, data on the large-scale distribution and structure of the diazotrophic communities across oceanic basins are sparse, as easily seen from the compilation of the available observations of diazotrophs and their distribution on global maps of the world's oceans (Luo et al. 2012). Vast regions of the oceans remain under-sampled both spatially and temporally with respect to diazotrophic phylotypes and abundances. In particular, the lack of observations is most noticeable in colder waters outside of the tropical oceans (Luo et al. 2012), probably because until recently marine BNF has been mainly investigated in warm waters where large blooms of *Trichodesmium* are easily noticed (Capone et al. 1997). In more recent years, *nifH*-based phylogenetic studies have firmly established the widespread distribution of diazotrophic microorganisms other than *Trichodesmium*, extending the distribution range of diazotrophs to the global oceans (Farnelid et al. 2011) and in particular to more temperate oceanic regions (Blais et al. 2012; Needoba et al. 2007).

Our study has focused on the detection of seven *nifH* phylotypes that have been previously identified as dominant in the SML (surface mixed layer) of the North Atlantic. The east–west transect we present here spans from 17°N to 40°N, latitudes that are under sampled with respect to diazotrophs. Previous east–west

transects crossed the Atlantic ocean at latitudes of 10°N and 0 – 20°N, omitting the North Atlantic gyre (Goebel et al. 2010; Langlois et al. 2008). Building on the work of Langlois et al. (2008) and Goebel et al. (2010) who presented results on some, but not all, of the same phylotypes discussed in our study, general similarities can be drawn between distribution patterns observed for these three transects. All show differential distributions of diazotrophs with UCYN A predominantly detected east of 40°W and *Hemiaulus-Richelia* symbiont mainly in the western Atlantic. Similarly to Langlois et al. (2008), the UCYN B and UCYN C phylotypes were also detected at higher abundances east of 40°W and the distributions of Gamma A and *Trichodesmium* were weakly correlated (R^2 = 0.39, Figure 2.4A and Figure 2.4C). Further south, *Trichodesmium* can also be abundant in the western Atlantic at the boundary between oligotrophic waters and Amazon River outflow (Subramaniam et al. 2008). Although not found in our study, Langlois et al. (2008) detected cluster III *nifH* phylotypes in the North Atlantic at higher latitudes than those sampled here.

2.4.2. Analysis of the community structure using multivariate statistics

2.4.2.1. East–west segregation

Our statistical analyses confirmed the observation that the diazotrophic phylotypes detected in our study inhabit primarily the SML (Figure 2.2) and hence we carried out the statistical analysis on the SML samples only. The geographically segregated east and west diazotrophic communities (confirmed by

ANOSIM; Table 2.1), were dominated by different phylotypes. The SIMPER analysis (Table 2.2) indicated that the *nifH* copies mL^{-1} of significant phylotypes contributing to the different community structure were unicellular cyanobacteria (UCYN A and UCYN B) and slightly higher abundances of *Trichodesmium* and Gamma A phylotypes east of 40°W, while the *Hemiaulus-Richelia* symbiont was the discriminant phylotype west of 40°W (Table 2.2). The diatom *Hemiaulus-Richelia* symbiont dominance near the North American coast, as estimated from the *nifH* abundances, was associated with much lower water temperature and higher nutrient concentrations, indicating a preference for these environmental conditions. Diatoms have a specific growth requirement for silicate, which could be a significant factor contributing to the observed distribution of the *Hemiaulus-Richelia* symbiont. The Amazon River is an important source of silicate for the western Atlantic and elevated Fe concentrations have been observed in higher silicate eddies from the Amazon River (Bergquist and Boyle 2006b). The influence of the Amazon River plume is felt in the warmer waters north of Brazil and in the Caribbean areas, where the *Hemiaulus-Richelia* association has often been detected (Goebel et al. 2010; Subramaniam et al. 2008; Foster et al. 2007; Carpenter et al. 1999). Therefore, the geographical distribution of the *Hemiaulus-Richelia* along the North American Eastern seaboard could also result from the transport of the diatom and its symbiont from the tropical Atlantic by eddy formation and spin-off from the Gulf Stream (Lee et al. 1991).

A high abundance of *Trichodesmium nifH* was measured in the surface sample in the eastern Atlantic at BATS (158 *nifH* copies mL^{-1}). Eight days before sampling at this station, tropical storm Sean passed BATS causing high winds and rainfall

(gusts of 91 kph and 0.05–0.12 in. rain at BATS). The resultant water column mixing and higher availability of nutrients may have contributed to higher *Trichodesmium nifH* abundances at that location.

2.4.2.2. Influence of aerosol concentrations on diazotroph distribution

High and low aerosol Fe concentration and aerosol origin superimposed on to the SML PCA plot (Figure 2.5) provided information on the relationships between high aerosol concentrations and specific *nifH* phylotypes present in different clusters. We assume here that measurements of high aerosol concentration equate to proportionally higher aerosol flux to surface waters. Eastern samples would therefore have presumably received high fluxes of mineral aerosols of North African/Saharan origin whereas western samples received low aerosol fluxes of either North American or Marine origin. As confirmed by ANOSIM (Table 2.1) and SIMPER (Table 2.2), and with the exception of *Hemiaulus-Richelia*, high mineral aerosol concentrations significantly correlated with diazotroph distribution in the SML. In addition, UCYN A, UCYN B, UCYN C, *Rhizosolenia-Richelia* and *Trichodesmium* were significantly associated with a North African/Saharan dust origin (Figure 2.5 and Supplemental Figure 3). This supports the prediction by Langlois et al. (2008) that high mineral aerosol concentrations would be correlated with high *Trichodesmium* and UCYN A copies mL^{-1} observed previously in the Cape Verde region, as well as general oceanic phytoplankton growth (Langlois et al. 2012; Franchy et al. 2013; Maranón et al. 2010; Langlois

et al. 2008; Duarte et al. 2006; Bonnet et al. 2005; Herut et al. 2005). Episodic dust storms deposit 10–50 g of dust m^{-2} to the eastern North Atlantic annually (Lawrence and Neff 2009) and hence supply this area with a variety of nutrients. Monthly averaged remote sensing data from MODIS over the time period of the research cruise (Supplemental Figure 4) shows high optical depth at 550 nm at the same stations that were recorded to have high mineral aerosol concentrations by Shelley et al. (2014). Precipitation data from TRMM (Supplemental Figure 4; Tropical Rainfall Measuring Mission; http://trmm.gsfc.nasa.gov/) showed that the ocean west of 40°W received high precipitation, again creating different environmental conditions in the North Atlantic east and west of 40°W during our study. The dust deposited on the ocean surface is a composite of many trace elements (Shelley et al. 2014; Buck et al. 2010; Baker et al. 2006; Viana et al. 2002; Goudie and Middleton 2001; Jickells 1999), macronutrients (Duce et al. 2008; Duarte et al. 2006; Guerzoni et al. 1999; Donaghay et al. 1991) and organic material (Wozniak et al. 2013; Mahowald et al. 2008) that may fulfill several important growth requirements for marine organisms, besides iron. Using Fe isotope ratios it was recently shown that Saharan dust is the dominant source of Fe to the North Atlantic, contributing 71 – 87% of all Fe (Conway and John 2014). The combination of high dust deposition, high temperatures and nitrate-limited surface waters near CVOO (Figure 2.3D) provided conditions favorable for diazotroph growth in the SML. These results support previous findings (Langlois et al. 2012; Rubin et al. 2011; Moore et al. 2009; Mills et al. 2004). However, our multivariate analyses including the suite of measurements from the GEOTRACES database, provides further statistical evidence that the high

abundance and diversity of diazotrophs in the eastern subtropical North Atlantic are linked to areas where surface waters receive high mineral dust deposition.

2.4.3. Presence of diazotrophic **nifH** *phylotypes below the surface mixed layer*

Although most of the high *nifH* copy numbers were found within the SML, the presence of a few phylotypes was also recorded below the SML. A section plot of chlorophyll fluorescence (Supplemental Figure 1) suggests that the diazotrophs detected at greater depth were close to or within the deep chlorophyll maximum, which occurs at the nutricline near the 1% light level. High *nifH* phylotype abundances below the deep chlorophyll maximum occurred primarily at the water mass boundaries between the sub-tropical gyre and oxygen depleted, high nutrient waters from the north west African upwelling (Supplemental Figure 1), which occurs mid-spring till mid-autumn at 15°N (Marcello et al. 2011). This OMZ reaches as far west as the Cape Verde Islands (minimum 40 mM at 400 m depth; Stramma et al. 2008), and supplies intermediate waters with macro-and micro-nutrients (Fitzsimmons et al. 2013; Rijkenberg et al. 2012; Bergquist and Boyle 2006a). The high dissolved Fe concentrations (41.5 nM) and higher phosphate concentrations (Figure 2.3D and Figure 2.3E and Supplemental Figure 1; Wurl et al. 2013; Zimmer and Cutter 2012) that were detected below the SML on the eastern side of the transect may provide optimal growth conditions for some of the diazotrophs found at the boundary of the gradient between the water masses. In particular, the Gamma A phylotype is thought to represent a heterotrophic

diazotroph and could possibly grow below the euphotic zone, but their physiology is not well understood to date. Sinking of the larger diazotrophs (*Trichodesmium* and diatom associated *Richelia*) has been reported before and this may account for the presence of *nifH* copies from these phylotypes below the euphotic zone (Scharek et al. 1999). Goebel et al. (2010) analysed *nifH* copies mL^{-1} down to a depth of 200 m and also found UCYN A and UCYN B down to a depth of 150 m specifically east of 40°W. However, our qPCR approach cannot differentiate between active growth or passive sinking.

2.4.4. Trace metals in the water column

Within the SML, correlation of *nifH* abundances with dissolved trace metals (e.g. Al, Mn, Ga, Ba, Pb) appeared to be related more to ocean circulation than to biological requirements. Some trace metal concentrations (e.g. Al, Ga, As) were higher in the Sargasso Sea than in the eastern basin close to the Saharan dust source, likely due to longer residence times in the oligotrophic gyre. Low nutrient concentrations in the Sargasso Sea result in lower productivity and hence lead to lower biological uptake and particle scavenging rates (Dammshäuser et al. 2011).

The standing stocks of trace metal concentrations only provide a snapshot at a given time and do not necessarily reflect the varying complex biological and physical mechanisms that control their concentrations. Nutrient draw down is often observed during phytoplankton blooms due to active nutrient uptake by the blooming species. Likewise, blooms of *Trichodesmium* have been shown to draw down dissolved iron and dissolved inorganic and organic phosphorus (Moore et

al. 2009). However, measurements of aerosol concentrations of trace metals are a better indicator of total atmospheric fluxes to the surface ocean and are more likely to be positively correlated with diazotroph abundance.

Dissolved Fe concentrations were patchy and generally high (in the SML 28 out of 37 stations had concentrations above 0.2 nM; Supplemental Figure 1), which may have contributed to the finding that dissolved Fe did not significantly influence diazotrophic distribution. It has been widely accepted that Fe influences diazotrophic distribution because of the high Fe requirements of the nitrogenase enzyme (Moore et al. 2009; Mills et al. 2004; Karl et al. 2002; Falkowski 1997). Fe uptake mechanisms are not completely understood and the preferred Fe source can differ greatly between organisms (Desai et al. 2012; Toulza et al. 2012). The comparison of different Fe sources with diazotrophic variation might explain some patterns seen in the data set. Another explanation could be that dFe is taken up rapidly by marine organisms leaving concentrations slow. Measurements of particulate Fe could be useful in evaluating this possibility.

Aerosol Al and Fe concentrations showed significant positive correlations with high diazotrophic abundances in the SML. Dissolved Al has historically been used as a tracer for dust deposition because it is a major component of mineral dust and was thought to be biologically inactive (Grand et al. 2014). However, recent studies suggests that variable dissolution of Al from wet and dry dust deposition as well as increased scavenging of Al in more productive ocean regions (Dammshäuser et al. 2011) will affect the usefulness of Al as a tracer for atmospheric Fe sources. Thus, dissolved Al concentrations in surface waters

may only be an accurate representation of dust deposition under specific conditions and the direct measurement of mineral concentrations will give a better idea of the fluxes supplied. In our study, where both dissolved and atmospheric data were available, aerosol Fe and Al concentrations were a better predictor of *nifH* phylotype abundances than dissolved Al concentrations in surface waters.

2.5. Conclusions

Basin wide *nifH* phylotype measurements from samples collected during US GEOTRACES cruises were used as a proxy to assess the large scale distribution patterns of several abundant marine diazotrophs found in the Atlantic Ocean. The west east transect spanned the North Atlantic from 10°W to70°W and 20°N to 40°N, from the surface down to 800 m. The distribution patterns of the *nifH* phylotypes showed that the communities on the eastern and western side of the Atlantic were significantly different. The western Atlantic diazotrophic community was characterized by the presence of *Hemiaulus-Richelia* association. In contrast, the eastern Atlantic diazotrophic community was dominated by the unicellular cyanobacteria groups (UCYN A, B and C), *Trichodesmium* and Gamma A. The eastern Atlantic community was associated with temperatures > 22°C in regions of high North African dust deposition, confirming the importance of aeolian dust deposition to the tropical eastern Atlantic ecosystem. Diazotroph abundance below the SML were associated with water masses with higher concentrations of remineralized nutrients, slightly enriched in PO_4^{3-} from either

the OMZ near the African Coast or the Gulf Stream on the western side of the Atlantic. Associations with other biologically relevant trace metals could not be conclusively demonstrated and dissolved Al concentrations could not be shown to predict the occurrence of *nifH* phylotypes.

2.6. Acknowledgements

We thank Robert Anderson, head of the US GEOTRACES program, for providing access to samples from the cruises, Chris Measures and Mariko Hatta, for providing dissolved Al and Fe concentrations and the calculation of the SML depths, and Douglas Wallace for discussion. We also thank Melanie Hammer for helping with the remote sensing data and Jennifer Tolman for helping with the data analysis and commenting on the manuscript. Funding for the nutrient and aerosol measurements was provided by the US National Science Foundation (grant OCE-0926092 to G. Cutter; grant OCE-0752351, 0929919 and 1132766 to W. M. Landing), and also *nifH* analysis by a Natural Science and Engineering Research Council of Canada RGPIN grant (45359) to Julie LaRoche. Jenni-Marie Ratten received financial support from the DAAD (Deutscher Akademischer Austausch Dienst), the SKS (Stiftung für Kanadastudien) and received a PhD stipend from the Transatlantic Ocean System Science and Technology (TOSST). Rebecca Langlois was supported by the Canadian Excellence Research Chair grant to D.W.R Wallace.

2.7. References

Altabet, M., 2006. Constraints on oceanic N balance/imbalance from sedimentary ^{15}N records. Biogeosciences Discussions, 3(4), pp.1121-1155.

Baker, A., Jickells, T., Witt, M., Linge, K., 2006. Trends in the solubility of iron, aluminium, manganese and phosphorus in aerosol collected over the Atlantic Ocean. Marine Chemistry, 98(1), pp.43-58.

Bergquist, B.A. and Boyle, E.A., 2006a. Dissolved iron in the tropical and subtropical Atlantic Ocean. Global Biogeochemical Cycles, 20(1).

Bergquist, B.A. and Boyle, E.A., 2006b. Iron isotopes in the Amazon River system: Weathering and transport signatures. Earth and Planetary Science Letters, 248(1), pp.54-68.

Blais, M., Tremblay, J., Jungblut, A. D., Gagnon, J., Martin, J., Thaler, M., Lovejoy, C., 2012. Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. Global Biogeochemical Cycles, 26(3), GB3022.

Bonnet, S., Guieu, C., Chiaverini, J., Ras, J., Stock, A., 2005. Effect of atmospheric nutrients on the autotrophic communities in a low nutrient, low chlorophyll system. Limnology and Oceanography, 50(6), pp.1810-1819.

Breitbarth, E., Oschlies, A., LaRoche, J., 2007. Physiological constraints on the global distribution of *Trichodesmium*–effect of temperature on diazotrophy. Biogeosciences, 4(1), pp.53-61.

Buck, C. S., Landing, W. M., Resing, J. A., Measures, C. I., 2010. The solubility and deposition of aerosol Fe and other trace elements in the North Atlantic Ocean: Observations from the A16N CLIVAR/CO_2 repeat hydrography section. Marine Chemistry, 120(1), pp.57-70.

Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., Carpenter, E. J., 1997. *Trichodesmium*, a globally significant marine cyanobacterium. Science, 276 (5316), pp.1221-1229.

Carpenter, E. J., Montoya, J. P., Burns, J., Mulholland, M. R., Subramaniam, A., Capone, D. G., 1999. Extensive bloom of a N_2-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. Marine Ecology Progress Series, 185, pp.273-283.

Church, M. J., Bjorkman, K. M., Karl, D. M., Saito, M. A., Zehr, J. P., 2008. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. Limnology and Oceanography, 53(1), p.63.

Church, M. J., Jenkins, B. D., Karl, D. M., Zehr, J. P., 2005. Vertical distributions of nitrogen-fixing phylotypes at stn ALOHA in the oligotrophic North Pacific Ocean. Aquatic Microbial Ecology, 38(1), pp.3-14.

Clarke K. R., Gorley R. N., 2006. PRIMER v6: User manual/tutorial. PRIMER-E, Plymouth.

Codispoti, L., 2006. An oceanic fixed nitrogen sink exceeding 400 tg N a^{-1} vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences Discussions, 3(4), pp.1203-1246.

Conway, T. M., John, S. G., 2014. Quantification of dissolved iron sources to the North Atlantic Ocean. Nature 511, pp.212–215.

Dammshäuser, A., Wagener, T., Croot, P. L., 2011. Surface water dissolved aluminum and titanium: Tracers for specific time scales of dust deposition to the Atlantic. Geophysical Research Letters, 38(24) L24601.

Desai, D.K., Desai, F.D. and LaRoche, J., 2012. Factors influencing the diversity of iron uptake systems in aquatic microorganisms. Environmental Bioinorganic Chemistry of Aquatic Microbial Organisms, p.103.

Donaghay, P. L., Liss, P. S., Duce, R. A., Kester, D. R., Hanson, A. K., Villareal, T., Gifford, D. J., 1991. The role of episodic atmospheric nutrient inputs in the chemical and biological dynamics of oceanic ecosystems. Oceanography, 4(2), pp.62-70.

Duarte, C. M., Dachs, J., Llabrés, M., Alonso-Laita, P., Gasol, J. M., Tovar-Sánchez, A., Agustí, S., 2006. Aerosol inputs enhance new production in the subtropical northeast atlantic. Journal of Geophysical Research, 111(G4), G04006.

Duce, R. A., LaRoche, J., Altieri, K., Arrigo, K. R., Baker, A. R., Capone, D. G., Zamora, L., 2008. Impacts of atmospheric anthropogenic nitrogen on the open ocean. Science, 320(5878), pp.893-897.

Falkowski, P. G., 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature, 387(6630), pp.272-275.

Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., Sørensen, S., Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PloS One, 6(4), e19223.

Fernández, A., Graña, R., Mouriño-Carballido, B., Bode, A., Varela, M., Domínguez-Yanes, J. F., Marañón, E., 2013. Community N_2 fixation and *Trichodesmium* spp. abundance along longitudinal gradients in the eastern subtropical North Atlantic. ICES Journal of Marine Science: Journal Du Conseil, 70(1), pp.223-231.

Fitzsimmons, J. N., Zhang, R., Boyle, E. A., 2013. Dissolved iron in the tropical north atlantic ocean. Marine Chemistry, 154, pp.87-99.

Foster, R., Subramaniam, A., Mahaffey, C., Carpenter, E., Capone, D., Zehr, J., 2007. Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical North Atlantic Ocean. Limnology and Oceanography, 52(2), pp.517-532.

Franchy, G., Ojeda, A., López-Cancio, J., Hernández-León, S., 2013. Plankton community response to Saharan dust fertilization in subtropical waters off the Canary Islands. Biogeosciences Discussions, 10(11), pp.17275-17307.

Gao, Y., Kaufman, Y., Tanre, D., Kolber, D., Falkowski, P., 2001. Seasonal distributions of aeolian iron fluxes to the global ocean. Geophysical Research Letters, 28(1), pp.29-32.

Goebel, N. L., Edwards, C. A., Church, M. J., Zehr, J. P., 2007. Modeled contributions of three types of diazotrophs to nitrogen fixation at station ALOHA. The ISME Journal, 1(7), pp.606-619.

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya, J. P., Edwards, C. A., Zehr, J. P., 2010. Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N_2 fixation in the tropical Atlantic Ocean. Environmental Microbiology, 12(12), pp.3272-3272.

Goudie, A., Middleton, N., 2001. Saharan dust storms: Nature and consequences. Earth-Science Reviews, 56(1), pp.179-204.

Grand, M.M., Buck, C. S., Landing W.M., Measures, C.I., Hatta, M., Hiscock, W.T., Brown, M., Resing. J.A., 2014. Quantifying the impact of atmospheric deposition on the biogeochemistry of Fe and Al in the upper ocean: A decade of collaboration with the US CLIVAR-CO2 Repeat Hydrography Program. Oceanography 27(1): pp.62–65.

Gruber, N., Sarmiento, J. L., 1997. Global patterns of marine nitrogen fixation and denitrification. Global Biogeochemical Cycles, 11(2), pp.235-266.

Guerzoni, S., Chester, R., Dulac, F., Herut, B., Loÿe-Pilot, M., Migon, C., Saydam, C., 1999. The role of atmospheric deposition in the biogeochemistry of the Mediterranean Sea. Progress in Oceanography, 44(1), pp.147-190.

Herut, B., Zohary, T., Krom, M., Mantoura, R. F. C., Pitta, P., Psarra, S., Frede Thingstad, T., 2005. Response of east Mediterranean surface water to Saharan dust: On-board microcosm experiment and field observations. Deep Sea Research Part II: Topical Studies in Oceanography, 52(22), pp.3024-3040.

Jickells, T., 1999. The inputs of dust derived elements to the Sargasso Sea; a synthesis. Marine Chemistry, 68(1), pp.5-14.

Jickells, T., An, Z., Andersen, K., Baker, A., Bergametti, G., Brooks, N., Torres, R., 2005. Global iron connections between desert dust, ocean biogeochemistry, and climate. Science, 308(5718), pp.67-71.

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D. and Stal, L., 2002. Dinitrogen fixation in the world's oceans. In The Nitrogen Cycle at Regional to Global Scales (pp. 47-98). Springer Netherlands.

Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., Foster, R. A., 2013. In situ identification and N_2 and C fixation rates of uncultivated cyanobacteria populations. Systematic and Applied Microbiology, 36(4), pp.259–271.

Langlois, R. J., Hümmer, D., LaRoche, J., 2008. Abundances and distributions of the dominant nifH phylotypes in the northern Atlantic Ocean. Applied and Environmental Microbiology, 74(6), pp.1922-1931.

Langlois, R. J., LaRoche, J., Raab, P. A., 2005. Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. Applied and Environmental Microbiology, 71(12), pp.7910-7919.

Langlois, R. J., Mills, M. M., Ridame, C., Croot, P., LaRoche, J., 2012. Diazotrophic bacteria respond to Saharan dust additions. Marine Ecology Progress Series, pp.470, 1-14.

Lawrence, C. R., Neff, J. C., 2009. The contemporary physical and chemical flux of aeolian dust: A synthesis of direct measurements of dust deposition. Chemical Geology, 267(1), pp.46-63.

Lee, T. N., Yoder, J. A., Atkinson, L. P., 1991. Gulf stream frontal eddy influence on productivity of the southeast US continental shelf. Journal of Geophysical Research: Oceans (1978–2012), 96(C12), pp.22191-22205.

Luo, Y., Doney, S., Anderson, L., Benavides, M., Bode, A., Bonnet, S., Carpenter, E., 2012. Database of diazotrophs in global ocean: Abundances, biomass and nitrogen fixation rates. Earth System Science Data Discussions, 5(1), pp.47-106.

Mahowald, N., Jickells, T.D., Baker, A.R., Artaxo, P., Benitez-Nelson, C.R., Bergametti, G., Bond, T.C., Chen, Y., Cohen, D.D., Herut, B. and Kubilay, N., 2008. Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. Global Biogeochemical Cycles, 22(4).

Maranón, E., Fernández, A., Mourino-Carballido, B., Martínez-García, S., Teira, E., Cermeno, P., Calvo-Díaz, A., 2010. Degree of oligotrophy controls the response of microbial plankton to Saharan dust. Limnology and Oceanography, 55(6), pp.2339-2352.

Marcello, J., Hernandez-Guerra, A., Eugenio, F., Fonte, A., 2011. Seasonal and temporal study of the northwest African upwelling system. International Journal of Remote Sensing, 32(7), pp.1843-1859.

Mills, M. M., Ridame, C., Davey, M., La Roche, J., Geider, R. J., 2004. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. Nature, 429(6989), pp.292-294.

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Zehr, J. P., 2010. Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. Science, 327(5972), pp.1512-1514.

Monteiro, F.M., Follows, M.J. and Dutkiewicz, S., 2010. Distribution of diverse nitrogen fixers in the global ocean. Global Biogeochemical Cycles, 24(3).

Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., Rijkenberg, M. J., 2009. Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. Nature Geoscience, 2(12), pp.867-871.

Needoba, J.A., Foster, R.A., Sakamoto, C., Zehr, J.P. and Johnson, K.S., 2007. Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. Limnology and Oceanography, 52(4), pp.1317-1327.

Noble, A.E., Lamborg, C.H., Ohnemus, D.C., Lam, P.J., Goepfert, T.J., Frame, C.H., Casciotti, K.L., DiTullio, G.R., Jennings, J.C. and Saito, M.A., 2012. Basin-scale inputs of cobalt, iron, and manganese from the Benguela-Angola front to the South Atlantic Ocean.

Ramakers, C., Ruijter, J. M., Deprez, R. H. L., Moorman, A. F., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters, 339(1), pp.62-66.

Redfield, A.C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton (pp. 176-192). James Johnstone memorial volume: university press of liverpool.

Rijkenberg, M.J., Steigenberger, S., Powell, C.F., Haren, H., Patey, M.D., Baker, A.R. and Achterberg, E.P., 2012. Fluxes and distribution of dissolved iron in the eastern (sub-) tropical North Atlantic Ocean. Global Biogeochemical Cycles, 26(3).

Rubin, M., Berman-Frank, I., Shaked, Y., 2011. Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. Nature Geoscience, 4, pp.529–534.

Scharek, R., Tupas, L.M. and Karl, D.M., 1999. Diatom fluxes to the deep sea in the oligotrophic North Pacific gyre at Station ALOHA. Marine Ecology Progress Series, 182, pp.55-67.

Schlitzer, R., 2012. Ocean data view.

Shelley, R.U., Morton, P.L. and Landing, W.M., 2015. Elemental ratios and enrichment factors in aerosols from the US-GEOTRACES North Atlantic transects. Deep Sea Research Part II: Topical Studies in Oceanography, 116, pp.262-272.

Shiller, A. M., 1997. Manganese in surface waters of the Atlantic Ocean. Geophysical Research Letters, 24(1), pp.1495-1498.

Sohm, J.A., Subramaniam, A., Gunderson, T.E., Carpenter, E.J. and Capone, D.G., 2011. Nitrogen fixation by *Trichodesmium* spp. and unicellular diazotrophs in the North Pacific Subtropical Gyre. Journal of Geophysical Research: Biogeosciences, 116(G3).

Stal, L. J., 2009. Is the distribution of nitrogen-fixing cyanobacteria in the oceans related to temperature? Environmental Microbiology, 11(7), pp.1632-1645.

Stramma, L., Johnson, G. C., Sprintall, J., Mohrholz, V., 2008. Expanding oxygen-minimum zones in the tropical oceans. Science, 320(5876), pp.655-658.

Subramaniam, A., Yager, P. L., Carpenter, E. J., Mahaffey, C., Bjorkman, K., Cooley, S., Capone, D. G., 2008. Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. Proceedings of the National Academy of Sciences of the United States of America, 105(30), pp.10460-10465.

Thompson, A. W., Zehr, J. P., 2013. Cellular interactions: Lessons from the nitrogen-fixing cyanobacteria. Journal of Phycology, 49(6), pp.1024-1035.

Toulza, E., Tagliabue, A., Blain, S. and Piganeau, G., 2012. Analysis of the global ocean sampling (GOS) project for trends in iron uptake by surface ocean microbes. PLoS One, 7(2), p.e30931.

Turk, K. A., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., Gilbert, J., 2011. Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. The ISME Journal, 5(7), pp.1201-1212.

Viana, M., Querol, X., Alastuey, A., Cuevas, E., Rodrıguez, S., 2002. Influence of African dust on the levels of atmospheric particulates in the Canary Islands air quality network. Atmospheric Environment, 36(38), pp.5861-5875.

Vitousek, P. M., Howarth, R. W., 1991. Nitrogen limitation on land and in the sea: How can it occur? Biogeochemistry, 13(2), pp.87-115.

Wozniak, A. S., Shelley, R. U., Sleighter, R. L., Abdulla, H. A. N., Morton, P. L., Landing, W. M. and Hatcher, P. G., 2013. Relationships among aerosol water soluble organic matter, iron and aluminum in European, North African, and Marine air masses from the 2010 US GEOTRACES cruise. Marine Chemistry, 154, pp.24-33.

Wurl, O., Zimmer, L., Cutter, G. A., 2013. Arsenic and phosphorus biogeochemistry in the ocean: Arsenic species as proxies for P-limitation. Limnology and. Oceanography, 58(2), pp.729-740.

Zehr, J. P., Mellon, M. T., Zani, S., 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes. Applied and Environmental Microbiology, 64(9), pp.3444-3450.

Zimmer, L.A. and G. A. Cutter., 2012. High resolution determination of nanomolar concentrations of dissolved reactive phosphate in ocean surface waters using long path liquid waveguide capillary cells (LWCC) and spectrometric detection. Limnology and Oceanography: Methods, 10, pp.568-580.

CHAPTER 3: THE DIVERSITY AND DISTRIBUTION OF *NIFH* PHYLOTYPES IN THE TROPICAL AND TEMPERATE NORTH ATLANTIC OCEAN

Jenni-Marie Ratten, Hannah Blanchard, Jaqueline Zorz, Jennifer Tolman, Dhwani Desai and Julie LaRoche

Contribution of authors:

Jenni-Marie Ratten:	Sample extraction and processing, data analysis and drafting of the manuscript
Hannah Blanchard:	Technical support with sample extraction, identification of *Candidatus* Atelocyanobacterium thalassa OTUs from the OTU table
Jaqueline Zorz:	Collection of Meteor 116 and AZMP HUD2014004 and HUD2014030 samples, technical support with sample extraction
Jennifer Tolman:	Technical support with qPCR assays and sequencing
Dhwani Desai	Support with bioinformatic processing
Julie LaRoche	Planning and discussion of the manuscript

3.0. Abstract

Until recently, the focus of research on marine diazotrophy has been in surface waters of tropical and subtropical oceanic regions dominated by cyanobacterial species. Recent observations in other oceanic regions suggest that non-cyanobacterial diazotrophic communities dominate in those areas. Using high-throughput Tag-sequencing, we recovered 2,809,851 *nifH* sequence reads from 407 samples collected throughout the Atlantic Ocean. At a clustering threshold of 96%, the pool of sequences yielded 2,655 operational taxonomic units (OTUs), among which 2195 represented novel OTUs. With a few exceptions, *nifH* sequence reads from tropical surface samples were dominated by *Trichodesmium spp.* and *Candidatus* Atelocyanobacterium thalassa with small fractions of reads from other clusters (18% relative abundance ± 21% relative abundance). We observed that clade 2 of *Candidatus* A. thalassa correlated with different haptophyte-derived chloroplast 16S rRNA gene sequences than clades 1 and 3. In contrast to the tropical ocean, samples north of 40°N and below the euphotic zone were dominated by *nifH* OTUs assigned to the proteobacterial dominated cluster I (sub-cluster Ip), cluster II and cluster III, although some *Candidatus* A. thalassa OTUs were also uncovered. Although the role of these non-cyanobacterial diazotrophic organisms in the nitrogen cycle is currently uncertain, the diversity and predominance of non-cyanobacterial diazotrophs observed in waters other than the tropical surface ocean suggest that these prokaryotes may contribute significantly to marine N_2 fixation, especially when considering that of water contained in the deep ocean provides a vast space

where non-cyanobacterial diazotrophs could thrive. Here we discuss the geographical range and the phylogenetic affiliation of some of the frequently recovered OTUs.

3.1. Introduction

Diazotrophs are a special group of bacteria and archaea that supply newly fixed nitrogen to the marine fixed nitrogen pool, counteracting the nitrogen-loss processes of anammox and denitrification, and essentially fertilizing vast areas of the open oceans that are beyond the influence of riverine or atmospheric fixed nitrogen input from land (Gruber 2008; Codispoti 2007; Hamersley et al. 2007; Codispoti et al. 2001; Ingall et al. 1994). Identification of novel diazotrophs is currently achieved by sequencing of the *nifH* gene, which codes for one of the subunits of the nitrogenase enzyme, necessary for the process of biological N_2 fixation. The earlier observations made by Farnelid et al. (2011) and supported by others (Fernandez et al. 2015; Loescher et al. 2014; Fernandez et al. 2011; Hewson et al. 2007; Metha et al. 2003) indicated that the cyanobacterial diazotrophs are dominant in the tropical and subtropical ocean surface waters while the diazotrophic communities at depth, in temperate regions and O_2-limited waters are dominated by non-cyanobacterial diazotrophs.

The recent technological advances in high-throughput sequencing technology have allowed unprecedented insights into microbial communities in various environments including the oceans (Sunagawa et al. 2015). Culture-independent methods hold the possibility to capture the full microbial diversity for specific

marker genes as well as the entire metabolic potential of a sample. In the exploration of diazotrophs in the marine environment, this approach has allowed for the dramatic expansion from capturing several hundreds of *nifH* sequences per study using Sanger sequencing to several thousand *nifH* sequences using high-throughput sequencing (Farnelid et al. 2011).

With a few exceptions in the Baltic Sea (Bentzon-Tilia et al. 2015; Severin et al. 2015; Farnelid et al. 2013), the Pacific (Cheung et al. 2016; Xiao et al. 2015) and a limited global survey at 10 locations (Farnelid et al. 2011), the reported *nifH* DNA sequences have originated primarily from Sanger sequencing of clone libraries from samples collected in the tropical waters of the North Atlantic and North Pacific Oceans (Luo et al. 2012; Bernavides et al. 2015). In this study, we applied high-throughput sequencing on the Illumina platform to assess the *nifH* diversity in samples collected from a range of sampling opportunities that included a time-series study in the Bedford Basin, a coastal ocean inlet on the Canadian east coast (Figure 3.1), as well as Atlantic research cruises. The transects spanned the Atlantic Ocean from 40°S to 60°N and 65 to 10°W. Our pool of *nifH* sequences (2,809,851 reads) originated from 407 stations and depths ranging from 1 – 1200 m and led to the determination of 2195 novel phylotypes of diazotrophs in addition to 460 that had previously been deposited to NCBI.

3.2. Methods

3.2.1. Sample Collection

Samples were collected during the Atlantic Zone Monitoring Programme (AZMP; HUD2014004 and HUD2014030) in spring and autumn 2014, the GEOVIDE cruise in summer 2014, the Meteor cruise M116 in summer 2015, US-GEOTRACES cruises Kn199 (winter 2010) and Kn204 (winter 2011), a section of the Polarstern ANTXXVI-I (November 2009), Discovery D361 (February 2011), and at four time points in the Bedford Basin through the Bedford Basin Monitoring Program on March 19, June 18, September 24 and December 17, 2014 at 1, 5, 10 and 60 m (Supplemental Table 3). Overall, sampling volumes varied between 0.28 and 6.2 L. AZMP HUD2014004 and HUD2014030 and Meteor M116 samples were prefiltered (160 µm mesh) onto 3 µm polycarbonate filters. All other samples were directly filtered onto 0.2 µm polycarbonate filters. Filters were either flash frozen in liquid nitrogen and stored at -80°C or immediately stored at -80°C until DNA extraction.

3.2.2. DNA Extraction

Except for the Meteor M116 cruise, DNA was extracted using the QIAGEN DNeasy Plant Mini Kit with a slightly modified cell lysis protocol as follows: filters were thawed at room temperature and incubated with 50 µL of lysozyme solution (5 mg mL^{-1} in TE buffer) for 5 min. 45 µL of ProteinaseK solution (20 mg mL^{-1} in MilliQ PCR grade water) and 400 µL of AP1 lysis buffer from the QIAGEN

DNeasy Plant Mini Kit were added and incubated for one hour at 52°C and 300 rpm on an orbital shaker. RNA was digested using 4 µL RNaseA from the QIAGEN DNeasy Plant Mini Kit at 65°C. Then, extraction continued according to the manufacturer's protocol, with a final elution volume of 50 µL. Concentrations and purity for DNA were determined using the NanoDrop 2000. Samples were stored in aliquots at 80°C until further analysis.

3.2.3. Meteor Nucleic Acid Extraction

DNA and RNA from samples collected during the Meteor M116 cruise were extracted using the Qiagen DNA/RNA AllPrep kit with a modified protocol: filters were thawed at room temperature and incubated for 5 min in 50 µL lysozyme solution (20 mg mL^{-1} in TE buffer). 50 µL Proteinase K and 600 µL RLT buffer with β-mercaptoethanol (10 µL per sample) were added and incubated at 52ºC and 300 rpm for 15 min. The lysate was filtered through a QIAshredder column by centrifuging the column at 14,000 rpm for 2 min. The supernatant was added onto a DNA spin column and extractions continued following the manufacturer's protocol. RNA and DNA samples were eluted in a final volume of 50 µL and aliquots were stored at -80 °C. DNA and RNA concentrations and purity were measured with a Nanodrop 2000. Samples were stored in aliquots at -80 °C until further analysis.

3.2.4. Illumina nifH *and 16S Library Preparation*

The presence of the *nifH* gene in each sample was tested using a nested PCR approach (Zehr and Turner 2001). The first amplification was carried out in 25 µL reactions containing 2.5 µL 10x buffer (Qiagen), 2 µL dNTPs (10 µM; Invitrogen), 4 µL $MgCl_2$ (25 mM; Qiagen), 2 µL each of *nifH* 3 and 4 primers (10 µM; Zehr et al. 2001), 0.3 µL BSA (20 mg mL^{-1}), 2.5 µL template, 0.125 µL Qiagen HotStar Taq polymerase (0.625 U) and 9.725 µL PCR graded water. PCR cycling conditions were 95°C for 15 min followed by 35 cycles of 95°C (1 min), 45°C (1 min), and 72°C (1 min), with a final 10 min at 72°C. The second PCR was performed with the same concentrations as the first except that the final reaction volume was 10 µL, $MgCl_2$ concentrations were reduced (1.2 µL of 25 mM stock), *nifH* 1/2 primers were used and 1 µL of template from the first reaction was added. PCR conditions were similar to the previous reaction, with the only modifications being an annealing temperature of 54°C instead of 95°C, and 28 PCR cycles instead of 35.

The first PCR step was repeated at a 1:10 template dilution for samples that tested positive for *nifH*. For sequencing on the Illumina MiSeq platform, PCR product from both first-round amplifications (*nifH* 3 and 4 primers) were combined and purified using a GeneJet PCR purification kit (Thermo Scientific). The second round of amplification was repeated with custom-made primers containing the *nifH* 1/2 primer sequence attached to the Illumina adaptor and bar code sequence for multiplexing in the Illumina MiSeq instrument (Supplemental Table 11). This amplification was carried out in 25 µL reactions with the same reagent

concentrations as above and an amplification protocol that was identical except for an annealing temperature of 52°C. From here on, sample preparation proceeded exactly as for the 16S rRNA gene sequencing described below.

The PCR-amplification and sequencing of the V6-V8 region of the bacterial 16S rRNA gene to identify chloroplast sequences, as well as the sequencing of *nifH* was performed at the Integrated Microbiome Resource (IMR) of the Centre for Comparative and Evolutionary Biology (CGEB) at Dalhousie University (Halifax, Canada). V6-V8 regions were amplified using custom 16S fusion primers; in addition to the universal primer sequences (B969F and BA1406R; Comeau et al. 2011), fusion primers contained Illumina adapters and barcodes for multiplexing at both ends of the fragments (Comeau et al. 2017). Amplifications were performed using two different dilutions (undiluted and 1:10). 25 µL reactions contained: 5 µL of 5xHF PCR Buffer, 0.5 µL dNTPs (40 mM), 5 µL forward and 5 µL reverse primer (1 µM), 0.25 µL Phusion polymerase (2 U μL^{-1}; Thermo Scientific), 2 µL template and 7.25 µL PCR-grade water. Cycling conditions were: 98°C (30 s), followed by 30 cycles of 98°C (10 s), 55°C (30 s) and 72°C (30 s). Final extension was performed for 4.5 min at 72 °C. The PCR product quality was verified using the E-gel 96-well high-throughput system (Invitrogen).

Amplification products of the V6-V8 regions and *nifH* were cleaned and normalized using the SequalPrep Normalization Plate Kit (Invitrogen). Samples were then multiplexed at equal volumes, quantified with the Qubit (Invitrogen) and loaded into the Illumina MiSeq platform as a 20 pM final denatured library according to manufacturer's instructions.

3.2.5. Bioinformatic Analysis

Raw Illumina paired-end reads of *nifH* and the 16S rRNA V6-V8 regions for chloroplast identification were preprocessed for QIIME (Quantitative Insights Into Microbial Ecology; Caporaso et al., 2010a) using the 16S amplicon analysis flow of the IMR (https://github.com/mlangill/microbiome_helper/wiki/16S-standard-operating-procedure; Comeau et al. 2017). Initial steps were run for both analyses. Reads were stitched together using PEAR (Paired-End reAd merger; Zhang et al. 2014). Reads that were short (*nifH* < 330 bp; 16S < 400 bp) low quality (Q<30 in >10% of samples), or uncertain (containing N) were removed using Comeau et al.'s (2017) pipeline. The splitting of 16S from *nifH* sequences was achieved through an in-house script using *nifH* primer sequences as the determining factor. Chimeric sequences were detected and removed using UCHIME (Edgar et al. 2011). The remaining reads were fed into the QIIME pipeline (Caporaso et al. 2010a). The QIIME open reference picking pipeline was run using SortMeRNA for closed reference picking against a curated *nifH* database or the Greengenes database for 16S rRNA sequences for chloroplast sequence identification and sumaclust for de novo OTU picking (*nifH* at 96% identity, 16S rRNA at 97% identity; Mercier et al. 2013; Kopylova et al. 2012; McDonald et al. 2012; Werner et al., 2012). The subsampling percentage was changed from 0.1% to 1.0%. PyNAST was used to perform the 16S rRNA alignments and MUSCLE for *nifH* alignments (Caporaso et al. 2010b; Edgar 2004). The last de-novo picking step, which is included in the pipeline by default, was suppressed due to extremely long processing time caused by the large

number of reads. Subsequently, singletons and low confidence OTUs were removed from the dataset. For 16S rRNA genes, taxonomies were assigned with RDP Classifier 2.2 (Wang et al. 2007). The samples of the *nifH* analysis were rarefied to 1500 reads per sample. OTUs that were assigned to chloroplast 16S rRNA were extracted from the 16S rRNA OTU table. This sub-table was rarefied to 1000 reads per sample.

3.2.6. Phylogenetic Analysis

To investigate the phylogenetic affiliation of the *nifH* OTUs obtained from the QIIME pipeline, reference genomes from 132 diazotrophs were selected from NCBI. Their *nifH* sequences were extracted and included in the alignment and phylogenetic analysis. The *nifH* sequences were trimmed to the position of nifH1 and nifH 2 primers. The alignment was constructed based on the protein sequence using MAFFT v. 7 (Yamada et al. 2016; Katoh et al. 2002) and returned to nucleotide sequences with PAL2NAL (Suyama et al. 2006). Ambiguous sequence alignment regions were removed using Gblocks, which reduced the number of alignment positions from 459 to 336 (Castresana 2000). The maximum likelihood analysis was performed using RAxML with the GTR-GAMMA model using default parameters and bootstrap values were calculated from 100 replicates (Stamakatis 2014). The complete tree (Supplemental Figure 7) was split into clusters and displayed with branch lengths showing the number of substitutions per site in iTOL (Letunic and Borg 2016).

A total of 35 reference OTUs displayed identity of at least 96% to *Candidatus* A. thalassa. Phylogenetic affiliations of these sequences, as well as sequences from *Trichodesmium erythraeum*, *Nostoc* sp. and *Crocosphaera watsonii*, were inferred as described above. The *Candidatus* A. thalassa sequences recovered from this study were classified according to the three clades previously identified by Thompson et al. (2014; Supplemental Figure 8).

16S rRNA gene sequences from chloroplasts that correlated with *Candidatus* A. thalassa *nifH* relative abundances were aligned using MAFFT v. 7, ambiguous sequence alignment regions were removed using Gblocks reducing alignment positions from 446 to 429, and maximum likelihood analysis was performed using RAxML with the GTR-GAMMA model using default parameters (Yamada et al. 2016; Stamakatis 2014; Katoh et al. 2002; Castresana 2000). Bootstrapping and displaying were performed as described above.

3.2.7. Statistical Analysis

Statistical analyses and creation of figures were performed in R version 3.2.1 using *ggplot2*, PRIMER-E version 6.1.12 and QIIME (R Core Team 2015; Wickham and Chang 2015; Caporaso et al., 2010a; Clarke and Gorley, 2006).

Rarefaction curves and alpha-diversity measures were created in QIIME using the alpha_rarefaction.py and alpha_diversity.py scripts (Caporaso et al., 2010a).

Beta-diversity was compared by Hellinger-transforming the relative abundance matrix and generating Bray-Curtis similarities using the R package *vegan* (Dixon,

2003). Bray-Curtis similarities were used to depict patterns of community composition over the year as well as in two-dimensional space using nonmetric multidimensional scaling (NMDS; Dixon, 2003). This was also done through *vegan*.

The single most common OTUs found in each sample were grouped in a histogram along both axes based on their presence or absence using *gplots* and displayed using *ggplot2*in the R package *gplots* (R core team 2015; Warnes et al. 2013).

The association between the *nifH* sequences of the three *Candidatus* A. thalassa clades with chloroplast 16S rRNA gene sequences were determined using SparCC, an algorithm that can find community interactions in high-throughput sequencing data more accurately than other approaches such as Pearson or Spearman correlation, because its algorithm is tailored to deal with the properties of genomic survey data (Friedman and Alm 2012).

3.3. Results

To explore spatial patterns of the diazotrophic community of the Atlantic Ocean comprehensively, we performed high-throughput sequencing of amplicons of a variable region of the *nifH* gene (ca. 360 bp, position in *nifH* gene ca 639 – 1000; Zehr et al. 2001) on 407 samples collected from 40°S to 60°N, 65°W to 10°W and 1 – 1200 m depth. These samples were taken during seven cruises (GEOTRACES: Kn199, Kn204 and GEOVIDE; AZMP: 2014 spring

(HUD2014004) and autumn (HUD2014030); Polarstern ANTXXVI-I and Meteor 116) and during the Bedford Basin Monitoring Program conducted by the Bedford Institute of Oceanography (Figure 3.1).

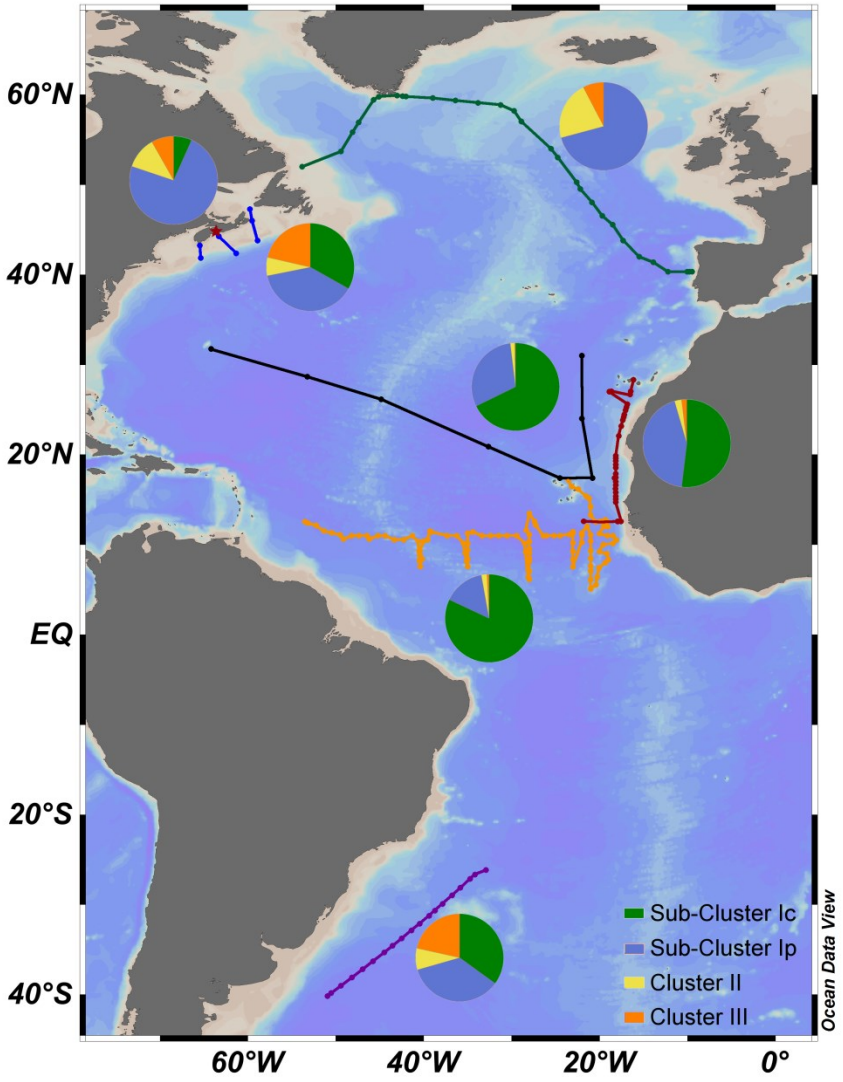


Figure 3.1: Sampling locations and relative abundance of the major clades of diazotrophs for each cruise throughout the Atlantic Ocean.

Samples were taken during the AZMP HUD2014004 and HUD2014030 cruises on the Scotian Shelf (blue), Discovery D361 on the Eastern North Atlantic (red), US-GEOTRACES Kn199 and Kn204 in the tropical Atlantic (black), GEOVIDE cruise from Portugal to Newfoundland (green), Meteor 116 in the tropical Atlantic (orange) and Polarstern ANTXXVI-I in the South Atlantic (purple) cruises as well as during the Bedford Basin Monitoring Program (red star). The relative abundance of the three major clusters of diazotrophs is indicated for each cruise by a pie chart (Zehr et al. 2003).

3.3.1. High-Throughput Sequencing of the **nifH** *gene*

After eliminating possible leak-through reads and removing sequences that included stop codons, the 407 DNA samples produced 2,809,851 *nifH* reads, with a maximum of 50,560 reads in one sample and a mean of 6887 reads. At 96% clustering identity, 2655 operational taxonomic units (OTUs) were identified, of which 545 had previously been deposited into the NCBI database, leaving 2195 novel OTUs that are first described in this study.

Samples were rarefied to a read number of 1500, which was chosen because this number showed a good saturation of total *nifH* diversity in most samples (Figure 3.2) while retaining a large proportion of overall samples (379 samples included in the analysis). The excluded 28 samples with lower read counts than 1500, originated mainly from high latitudes where *nifH* diversity is low.

On average, the total richness in samples reached saturation at 3500 reads/samples, which corresponded on average to the detection of 85 OTUs (Figure 3.2). However, saturation varied greatly among samples depending on sampling location. Open ocean samples collected at latitudes higher than 40°N were the least diverse. In the high latitude samples, the entire community composition was captured at 1500 reads with the detection of an average of 9.4 ± 4.8 OTUs (61 samples: 2.3 – 21.6 OTUs detected in samples at latitudes higher than 40°N). Open ocean samples collected between 40°S and 40°N spanned the range of 5.2 to 190.8 OTUs at 5000 reads (average = 96.4 ± 43.2), whereas samples from the Bedford Basin had not yet reached complete saturation at 5000 reads. Coastal samples from the AZMP HUD2014004 and HUD2014030 cruises

on the Scotian Shelf and the Bedford Basin displayed the highest OTU richness, with an average of 89.7 ± 76.1 OTUs detected (49.9 – 340.2 OTUs). Most unique OTUs were found in the Bedford Basin in December at 1 and 10 m depth (340.2 and 291 OTUs respectively; Figure 3.2).

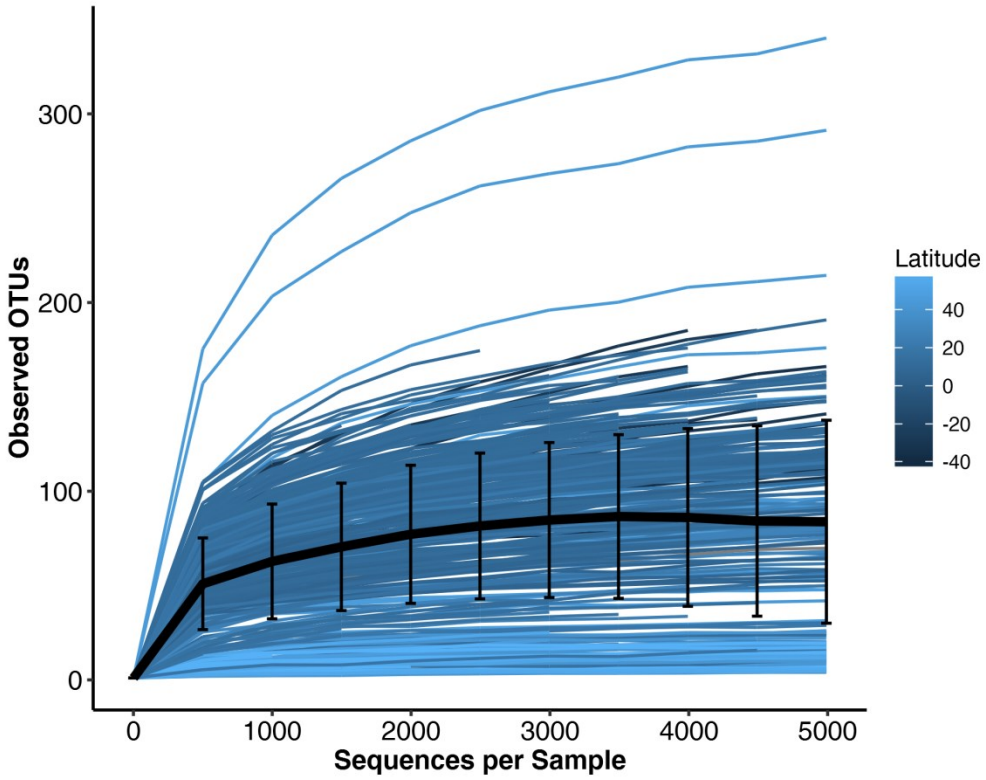


Figure 3.2: Rarefaction curve of recovered *nifH* OTUs.

The number of recovered OTUs is shown at increasing sequencing depths for every sample. Samples are coloured according to latitude, as indicated by the legend.

Similarly to OTU richness, alpha-diversity measures varied with latitude (Figure 3.3). Alpha-diversity also changed with depth. Figure 3.3 depicts the overall diversity in samples (Shannon diversity) and the total species estimate (Chao1).

Highest overall Shannon diversity was observed in surface samples (an average of 4.96 with a maximum and minimum of 7.02 and 1.55 in the Bedford Basin on Dec. 17th 2014 at 1 m and during the Meteor 116 cruise at 1 – 19 m, 21.00°W, 14.20°N respectively; Figure 3.3A). Shannon diversity index decreased with depth to an average of 2.82 and 2.28 at depths of 20 – 200 m and > 200 m, respectively (Figure 3.3A). Lowest overall diversity was observed in samples north of 50°N.

The richness estimate Chao1 followed a similar trend to the Shannon diversity (Figure 3.3B). Samples that contained the highest estimated number of OTUs were surface samples from 40°S to 20°N (on average 122 OTUs with a maximum and minimum of 271 and 32). OTU richness decreased with increasing latitude after 20°N and with depth. Depths of 20 – 200 m and > 200 m averaged an estimate of 55 and 38 OTUs, respectively (Figure 3.3B).

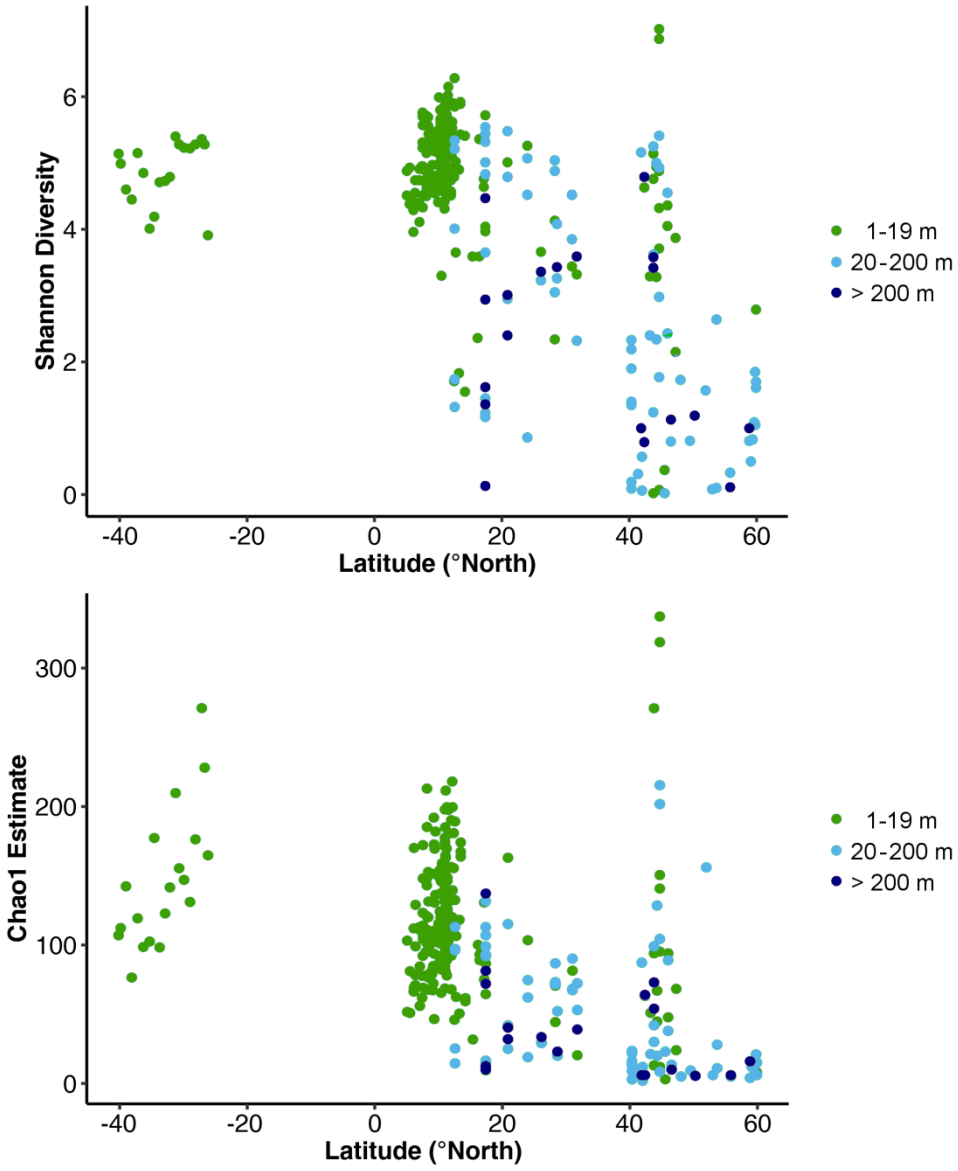


Figure 3.3: Alpha-diversity along latitude and depth.

The Shannon diversity (A) and richness estimator Chao1 (B) were calculated for each sample. Diversity measures are plotted along latitude and divided into three depths (1 - 19 m, 20 – 200 m and > 200 m).

3.3.2. Phylogenetic affiliation of diazotrophs in the Atlantic Ocean

Our opportunistic sampling scheme, although extensive, resulted in a predominance of samples originating from the tropical North Atlantic (61% from 0-20°N, Table 3.1). Because distribution patterns of diazotroph clusters appear to vary with latitude, with cyanobacteria dominating at lower latitude, this carries implications for the detection frequency of *nifH* reads from the non-cyanobacterial clusters (Figure 3.4).

Table 3.1: Sample numbers for oceanic regions and depths.

Oceanic region	Number of samples	%	1 - 19 m	%	20 - 200 m	%	> 200 m	%
20 – 40°S	19	5	19	5	0	0	0	0
0 – 20°N	249	61	230	57	13	3	6	1
20 – 40°N	44	11	19	5	17	4	8	2
40 – 60°N	95	23	29	7	48	12	18	4
Total:	**407**	100	**297**	73	**78**	19	**32**	8

For visualisation purposes, the large overall tree was divided into four sub-trees according to three main clusters (I, II and III), with cluster I being further divided into cluster Ic that contained the cyanobacterial reference sequences, and cluster Ip that contained mostly proteobacterial reference sequences (Figure 3.4). Greatest diversity was found in sub-cluster Ip, which contained over half of the identified OTUs (53%). Cluster III, cluster III and sub-cluster Ic comprised 26, 11 and 10% of OTUs, respectively (Supplemental Figure 7).

The clusters varied in relative abundances with depth (Figure 3.4). As expected, cyanobacteria (sub-cluster Ic) were the only group predominantly found in the surface and decreasingly at depth. Cluster II, cluster III and sub-cluster Ip were most abundant between 20 – 200 m. Cluster III was the most evenly distributed throughout all depths.

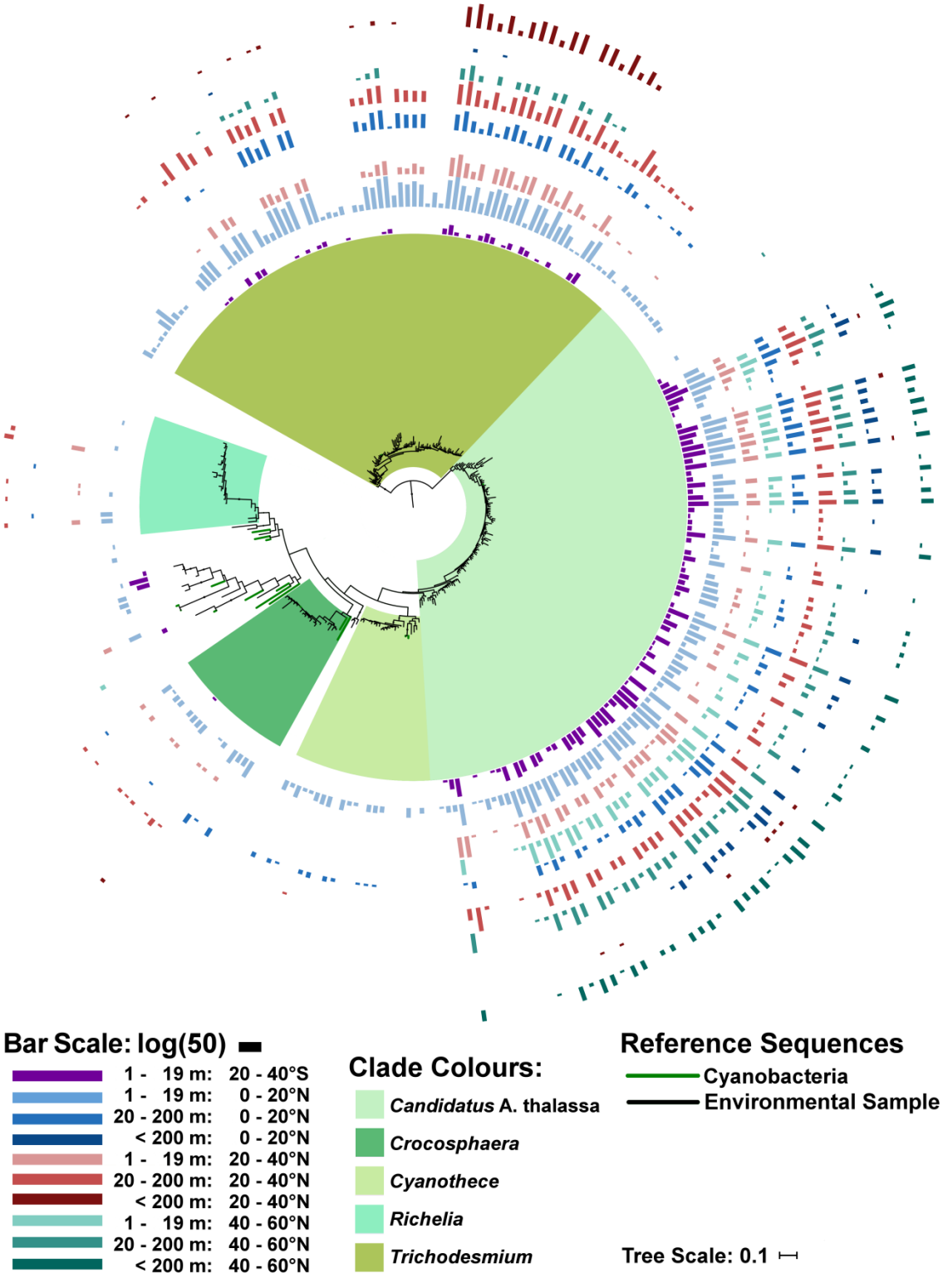

Bar Scale: log(50)
1 - 19 m: 20 - 40°S
1 - 19 m: 0 - 20°N
20 - 200 m: 0 - 20°N
< 200 m: 0 - 20°N
1 - 19 m: 20 - 40°N
20 - 200 m: 20 - 40°N
< 200 m: 20 - 40°N
1 - 19 m: 40 - 60°N
20 - 200 m: 40 - 60°N
< 200 m: 40 - 60°N
Clade Colours:
Candidatus A. thalassa
Crocosphaera
Cyanothece
Richelia
Trichodesmium
Reference Sequences
Cyanobacteria
Environmental Sample
Tree Scale: 0.1

Figure 3.4A: Phylogenetic tree and relative abundance data for sub-cluster Ic of cyanobacterium-related *nifH* sequences recovered throughout the Atlantic Ocean.

Of the 2655 OTUs, 267 were phylogenetically assigned to sub-cluster Ic. The phylogenetic affiliation of these OTUs and reference sequences was inferred by maximum likelihood tree building based on the GTR-GAMMA model of codon-aligned *nifH* sequences (MAFFT v. 7; Yamada et al. 2016; Stamakatis 2014; Katoh et al. 2002). Bootstrap support from 100 replicates are displayed as circles when > 50. The tree was displayed with branch lengths showing the number of substitutions per site. Leaves of sequences from reference genomes are coloured according to their taxonomy, while black leaves indicate environmental sequences. The outer bars indicate the logarithmically transformed number of sequences present for each OTU in the OTU table rarefied to 1500 reads. Bars are coloured based on the location (South Atlantic Ocean, tropical, mid-latitude and temperate North Atlantic Ocean) and depth of their detection (1 – 19 m, 20 – 200 m, > 200 m).

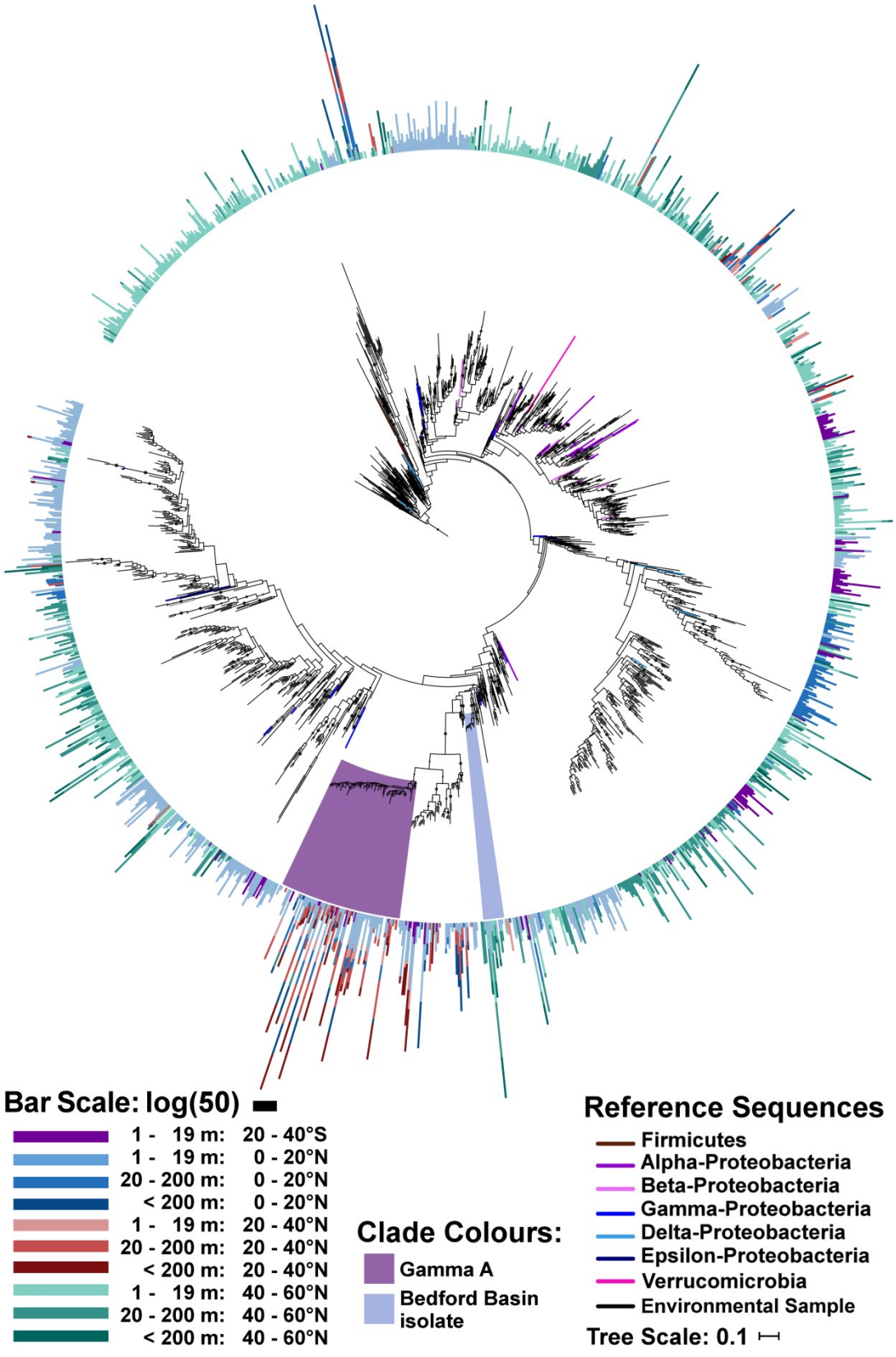

Bar Scale: log(50)
1 - 19 m: 20 - 40°S
1 - 19 m: 0 - 20°N
20 - 200 m: 0 - 20°N
< 200 m: 0 - 20°N
1 - 19 m: 20 - 40°N
20 - 200 m: 20 - 40°N
< 200 m: 20 - 40°N
1 - 19 m: 40 - 60°N
20 - 200 m: 40 - 60°N
< 200 m: 40 - 60°N
Clade Colours:
Gamma A
Bedford Basin isolate
Reference Sequences
Firmicutes
Alpha-Proteobacteria
Beta-Proteobacteria
Gamma-Proteobacteria
Delta-Proteobacteria
Epsilon-Proteobacteria
Verrucomicrobia
Environmental Sample
Tree Scale: 0.1

Figure 3.4B: Phylogenetic tree and relative abundance data for sub-cluster Ip of proteobacterium-related *nifH* sequences recovered throughout the Atlantic Ocean.

Of the 2655 OTUs, 1407 were phylogenetically assigned to sub-cluster Ip. Parameters for the tree construction were the same as in Figure 3.4A. Stacked bar chart represent the relative abundance of each OTU in the selected regions. Black leaves represent environmental sequences as in Figure 3.4A. The purple clade indicates sequences related to the heterotrophic diazotroph isolated from the Bedford Basin (Chapter 5).

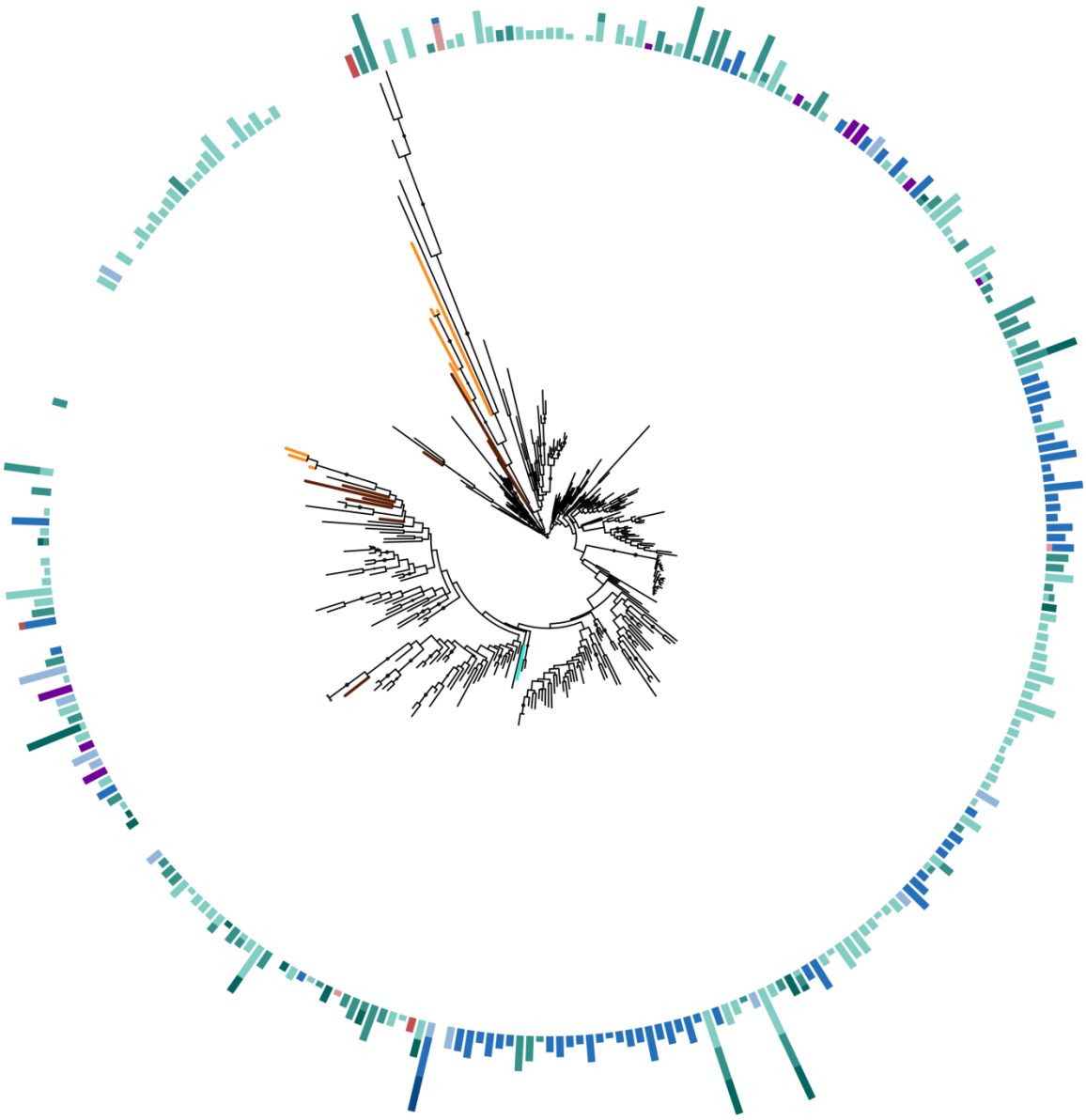

Bar Scale: log(50)
1 - 19 m: 20 - 40°S
1 - 19 m: 0 - 20°N
20 - 200 m: 0 - 20°N
< 200 m: 0 - 20°N
1 - 19 m: 20 - 40°N
20 - 200 m: 20 - 40°N
< 200 m: 20 - 40°N
1 - 19 m: 40 - 60°N
20 - 200 m: 40 - 60°N
< 200 m: 40 - 60°N
Reference Sequences
Archea
Firmicutes
Chlorobia
Environmental Sample
Tree Scale: 0.1

Figure 3.4C: Phylogenetic tree of cluster II *nifH* sequences recovered throughout the tropical and temperate Atlantic Ocean.

Of the 2655 OTUs, 292 aligned in cluster II. Parameters for the tree construction were the same as in Figure 3.4A. Stacked bar chart represent the relative abundance of each OTU in the selected regions. Black leaves represent environmental sequences as in Figure 3.4A.

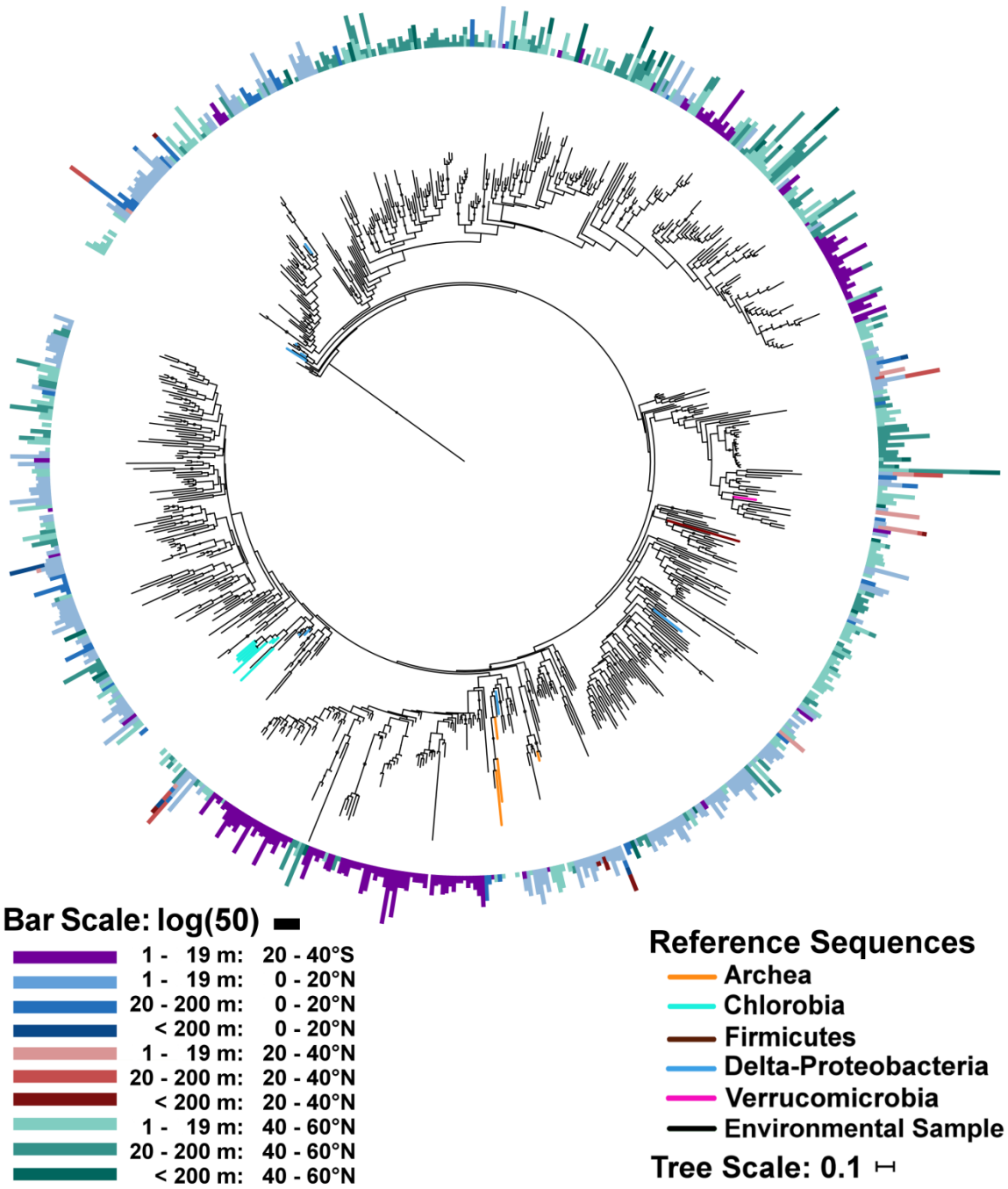


Figure 3.4D: Phylogenetic tree and relative abundance data for cluster III *nifH* sequences recovered throughout the Atlantic Ocean.

Of the 2655, 690 aligned in cluster III. Parameters for the tree construction were the same as in Figure 3.4A. Stacked bar chart represent the relative abundance of each OTU in the selected regions. Black leaves represent environmental sequences as in Figure 3.4A.

Overall, sub-cluster Ic OTUs were distributed through most samples and accounted for the highest number of reads (Figure 3.5). In the rarefied OTU table, the most frequently detected OTU was DQ530493 (32948 reads), for which the closest reference genomic sequence is *Trichodesmium erythraeum* IMS101 (identity: 97%), which can be explained by the predominance of samples from tropical waters. This was followed by sub-cluster Ip, cluster II and lastly, cluster III. Of the 22 most common sequences (those with >6000 cumulative reads), 16 were from sub-cluster Ic, while 5 and 1 were from sub-cluster Ip and cluster II respectively. All those from sub-cluster Ic were ≥ 96% identical to a sequence from a reference genome, while none of the others were (Figure 3.5).

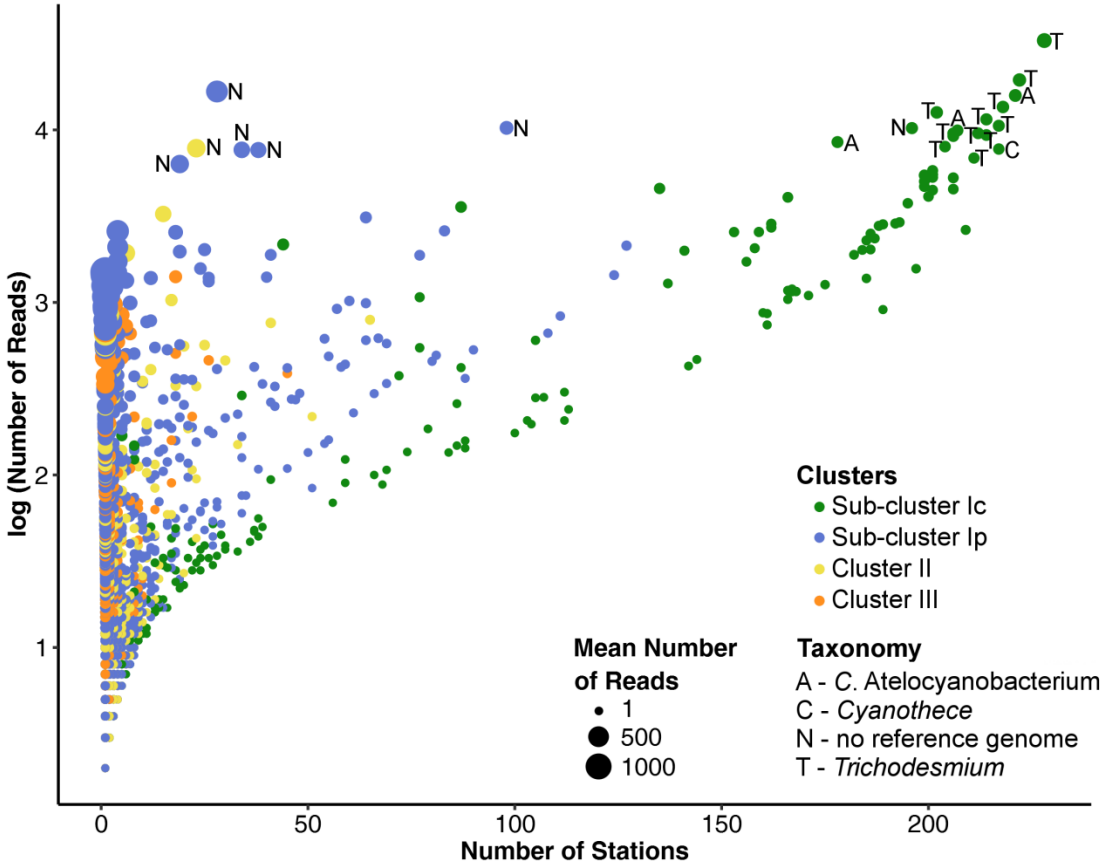


Figure 3.5: Abundances and distribution of *nifH* OTUs.

For each OTU, the number of reads was extracted from the rarefied OTU table and the logarithmically transformed number was plotted against the number of stations in which this OTU was found. The size of the dot indicates the mean number of reads a specific OTU displayed in the samples in which it occurred. For total reads of the OTU > 6000, the closest reference genome is indicated by A = *Candidatus* A. thalassa, C = *Cyanothece*, T – *Trichodesmium* and no reference genome found within 96% sequence identity = N.

3.3.3. Community Structure

The comparison of the diazotrophic communities based on Bray-Curtis similarity indicated that samples mainly varied in composition along latitude (Figure 3.6). The communities north of 40°N clustered separately from all other samples, regardless of depth (Figure 3.6A). A higher resolution of the high-density cluster

shows that there are three sub-clusters, one containing samples mainly from -20 to -40 and 20 to 40°N, one from samples at sampling depth of 20 m and greater. The largest cluster including surface samples from 0 – 20°N (Figure 3.6B).

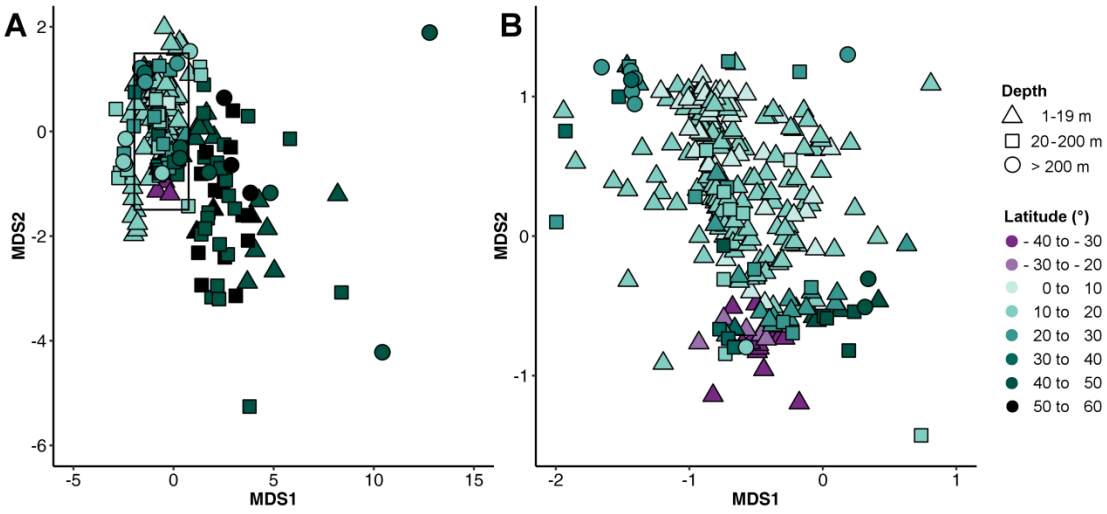


Figure 3.6: Diazotrophic community similarity throughout the Atlantic Ocean. Non-linear Multi-Dimensional Scaling (NMDS) was used to plot sample similarity according to their taxonomic composition and abundance. (A) includes all but 3 outliers from the GEOVIDE cruise transect, whereas (B) shows a higher resolution of the crowded samples within the black box in A. The *nifH* counts were Hellinger transformed and NMDS plots were created based on Bray-Curtis dissimilarity between samples. Each point represents one sample and is colour-coded according to latitude. The shape of the symbol represents the depth of collection.

The analysis of relative abundances of each *nifH* cluster and the most common OTU in each sample indicates similar patterns (Figure 3.7 and Supplemental Figure 5). Cyanobacterial proportions were highest in tropical surface samples with occasional spikes in proteobacteria. An increase in the relative abundance of cluster II and III sequences in the surface was seen north of 40°N and south of

20°S (Supplemental Figure 5). In the 20 – 200 m depth section, clusters varied more dramatically with samples dominated by cyanobacterial, proteobacterial or cluster II sequences (Supplemental Figure 5). This is also the case for samples collected below 200 m (Supplemental Figure 5).

The most abundant OTU was assigned to the major *nifH* clusters and compared to the reference genome database in NCBI using BLAST search (Altschul et al. 1990; Figure 3.7 and Supplemental Figure 6). Identity to reference genomes varied from 99% to 75%. Reference genomes with > 96% identity are noted in Figure 3.7. The most frequently identified reference genomes were *Trichodesmium erythraeum* (152 times, identity 93 – 99%) and *Candidatus* A. thalassa (75 times, identity 95 – 99%). The next most common and most abundant OTU had 78% sequence identity with *Magnetococcus marinus* (found 19 times). Together, this described 65% of samples. Clustering samples based on presence/absence of the most common OTUs resulted in three major clusters (L1 – L3; Figure 3.7; Warnes et al 2013). Latitude was associated with these clusters. L1 contained mainly samples collected in the South Atlantic and above 40°N with a dominant presence of *Candidatus* A. thalassa and cluster II sequences. L2 mainly contained samples from 20 - 60°N with many sub-cluster Ip OTUs that occurred in only one sample (56 OTUs) and the largest cluster (L3) was made up of mainly tropical samples dominated by *Trichodesmium* and *Candidatus* A. thalassa and some *nifH* sequences of sub-cluster Ip (0 - 20°N; Figure 3.7). Among these sub-cluster Ip OTUs was the OTU with 78 – 80% identity to *Magnetococcus marinus*. Also abundant was an OTU with 88% identity

to *Confluentimicrobium* sp. Both OTUs were found between 5 and 17°N and had previously been identified in the tropical east Atlantic, the northern South China Sea (Turk et al. 2011; Kong et al. 2011), or only the eastern tropical Atlantic, respectively (Turk et al. 2011). In the southern transect, an OTU of cluster III that showed 81% identity to *Desulfovibrio alkalitolerans* was found to dominate 6 of 12 samples. It was not previously deposited in the database. The closest NCBI sequence showed 92% identity and was isolated from a salt marsh (Gamble et al. 2010).

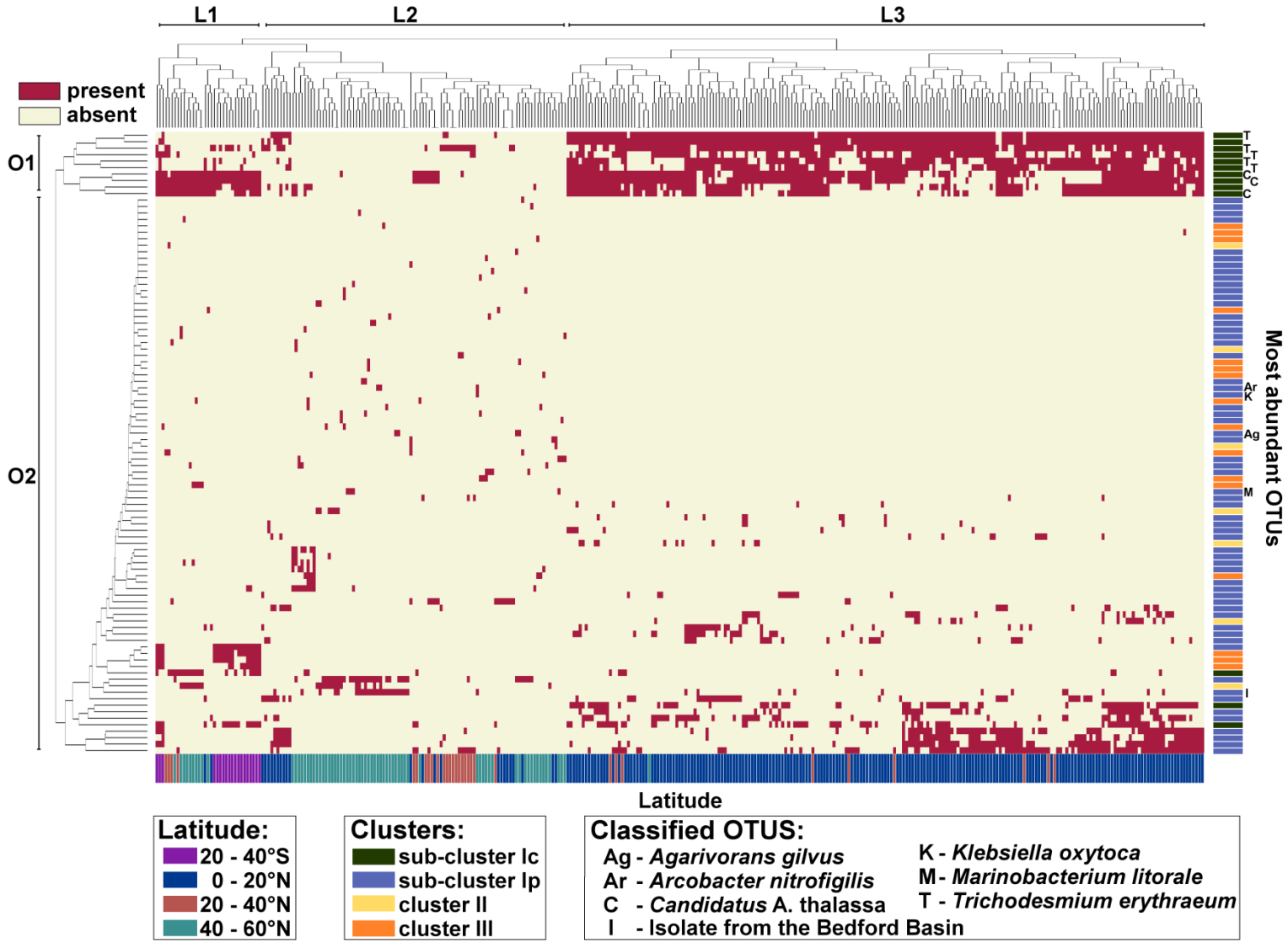

L1
L2
L3
present
absent
O1
O2
Most abundant OTUs
Latitude
Latitude:
20 - 40°S
0 - 20°N
20 - 40°N
40 - 60°N
Clusters:
sub-cluster Ic
sub-cluster Ip
cluster II
cluster III
Classified OTUS:
Ag - Agarivorans gilvus
Ar - Arcobacter nitrofigilis
C - Candidatus A. thalassa
I - Isolate from the Bedford Basin
K - Klebsiella oxytoca
M - Marinobacterium litorale
T - Trichodesmium erythraeum

Figure 3.7: Cluster-assignment and distribution of the most abundant *nifH* OTUs in each sample.

The single most abundant *nifH* OTUs were extracted from each sample, assigned to a major cluster (sub-cluster Ic – green, sub-cluster Ip – blue, cluster II – yellow, cluster III – orange), and compared against the reference genome database in NCBI using BLAST search (Altschul et al. 1990). Reference genomes for sequences of at least 96% identity are indicated on the y-axis. Samples were clustered based on presence/absence in each sample using *gplots* in R (Warnes 2013). The horizontal axis is colour-coded according to latitude (20 - 40°S – magenta, 0 - 20°N – dark blue, 20 - 40°N – red, 40 - 60°N – turquoise) while the vertical axis represent the most abundant taxa in each sample coloured by cluster.

As for the clustering of reads at the OTU level, the identity cutoff of 96% was used to compare and assign the *nifH* OTUs to reference genomes of sequenced diazotrophs. Based on this identity cutoff, dominant OTUs from 207 (54.6%) samples could be classified (Supplemental Figure 6). All but six sequences with a 96% or higher identity were assigned to *Trichodesmium erythraeum* or *Candidatus* A. thalassa. The other six were assigned to *Agarivorans gilvus* (twice), *Arcobacter nitrofigilis*, *Klebsiella oxytoca* and *Marinobacterium litorale* (twice; Figure 3.7). In contrast to the cyanobacteria, *nifH* OTUs that could be assigned to known genera of other clusters (6 of 73 sequences) with some confidence (i.e. > 96% identity) were all derived from samples north of 45°N. Since *Trichodesmium erythraeum* and *Candidatus* A. thalassa dominated in the tropical ocean, samples collected there were classified with more certainty compared to samples further north or south (Supplemental Figure 6). The classified cyanobacteria of the southern transect were exclusively *Candidatus* A. thalassa.

3.3.4. *Candidatus* Atelocyanobacterium thalassa

SparCC correlation was used to explore possible symbiotic relationships between *Candidatus* A. thalassa clades I – III (Supplemental Figure 8) and its potential hosts. For this, all OTUs clustering within the *Candidatus* A. thalassa clade were extracted (35 OTUs) and were assigned to sub-clades 1-3 through a phylogenetic analysis that included reference sequences from Thompson et al. (2014; Supplemental Figure 8). These OTUs were correlated with chloroplast 16S rRNA

gene abundances from the same samples (the phylogeny of chloroplast 16S rDNA is shown in Supplemental Figure 9). The strength of significant positive correlations ($p \leq 0.05$) is recorded in Table 3.2. Sequences of clades 1 and 3 consistently correlated with Haptophyceae New Reference OTU5 (Hapt NROTU5), whereas clade 2 correlated with Haptophyceae New Reference OTU 32 (Hapt NROTU32; Figure 3.8). Both these chloroplast 16S rRNA OTUs show 99% identity with *Braarudosphaera bigelowii* when BLASTed against the NCBI database (Altschul et al. 1990). Clade 1 also shows strong correlation with Stramenopiles OTU 137837 (Stram 137837; Table 3.2), which was identified as a member of Chrysophyceae when compared to the phytoref database (Decelle et al. 2015).

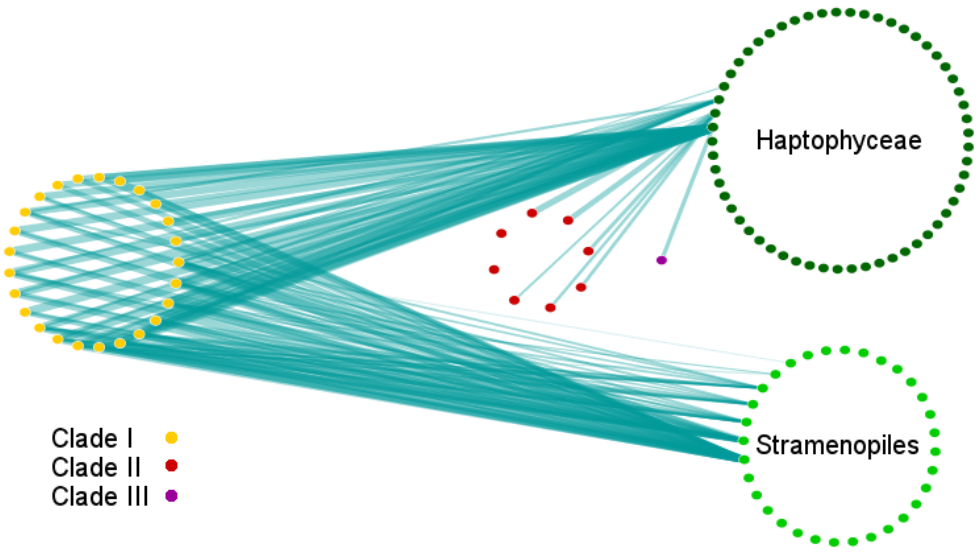


Figure 3.8: Association Network of *Candidatus* A. thalassa OTUs of clades 1, 2 and 3 with chloroplast 16S rRNA gene OTUs.

SPARCC was used to establish correlations between OTUs of *Candidatus* A. thalassa clades and OTUs of 16S rRNA genes from chloroplasts. Only significant positive correlations are shown, with edge thickness representing the strength of the correlation.

Table 3.2: SparCC correlations of *Candidatus* A. thalassa clades I, II and III with chloroplast 16S rRNA genes.

Clade[1)]	GI number[2)]	Hapt.[3)] 233100	Hapt. OTU5 B. b.[4)]	Hapt. OTU637	Hapt. OTU32 B. b.	Stram.[5)] 1106681	Stram. 137837	Stram. 354695	Stram. 592855	Stram. 746239	Stram. OTU102	Stram. OTU153
	CP001842	0.28[6)]				0.25						
	EU187510	0.36				0.27		0.18				
	EU187520	0.4	0.2		0.2	0.34	0.18	0.27	0.19			
	EU187521	0.4	0.18		0.17	0.32		0.24	0.17			
	EU187527	0.35				0.27		0.17				
	EU187528	0.28				0.24						
	EU187529	0.27				0.23						
	EU187530	0.41	0.19		0.17	0.33	0.18	0.27	0.18		0.15	
	EU187535	0.39	0.18			0.31	0.16	0.23	0.17			
	EU187538	0.4	0.2		0.19	0.33	0.18	0.28	0.2		0.16	
	EU187539	0.35	0.19			0.32		0.21	0.17			
	EU187542	0.41	0.18		0.17	0.32	0.17	0.25	0.18			
1	EU187544	0.37	0.2		0.17	0.33	0.17	0.23	0.18			
	EU187546	0.31				0.27						
	EU187547	0.25				0.24						
	EU187549	0.36	0.19		0.17	0.33	0.18	0.24	0.2			
	EU187550	0.25				0.21						
	EU187558	0.34				0.29		0.18				
	EU187560	0.22				0.21						
	EU187563	0.28				0.26						
	EU187564	0.36	0.2			0.31		0.18				
	EU187567	0.15	0.46	0.2	0.13	0.18	0.34	0.17	0.3	0.19	0.14	
	EU187571	0.38	0.19		0.18	0.32	0.17	0.24	0.17			
	EU187572	0.32				0.25		0.16				
	HQ456060	0.33	0.19			0.27	0.18	0.18				

2	EF568479		0.2
	EF568480		0.18
	EF568483		0.3
	EF568484		0.22
	EF568540		0.21
	HQ455985		0.27
3	HM210392	0.23	

1) The *Candidatus* A. thalassa clade that the specific *nifH* sequence aligns with (Thompson et al. 2014).
2) GI numbers of *Candidatus* A. thalassa sequences found during this study.
3) Hapt. = Haptophyceae
4) B. b. = *Braarudosphaera bigelowii*
5) Stram. = Stramenopiles
6) Each value shows the strength of a correlation between the abundances of *nifH* sequence and chloroplast 16S rRNA sequence. Only correlations with $p \leq 0.05$ are shown.

3.4. Discussion

Our extensive *nifH* sequence dataset collected throughout the pelagic Atlantic Ocean, has demonstrated the high diversity and wide geographic distribution of *nifH*, a functional gene for diazotrophs. With a few exceptions (Farnelid et al. 2011, Halm et al, 2012), pelagic studies of marine N_2 fixation have focussed on tropical surface oceans (Luo et al. 2012), which has greatly underestimated diversity and distribution of diazotrophs. Snapshots into other regions, including the temperate, deep and O_2 depleted ocean, have shown that diazotrophic communities in those regions are dominated by heterotrophic or chemoautotrophic microorganisms rather than the photoautotrophic cyanobacteria found in the tropical surface ocean (Loescher et al. 2014; Farnelid et al. 2011; Fernandez et al. 2011; Metha et al. 2003). Since, the sequencing depth achieved by high-throughput technologies, has supported the hypothesis that heterotrophic diazotrophs make up a large part of many N_2 fixing communities throughout the oceans (Cheung et al. 2016; Bentzon-Tilia et al. 2015; Severin et al. 2015; Xiao et al. 2015; Farnelid et al. 2013).

3.4.1. The diazotrophic community of the Atlantic Ocean

In this study, we obtained 2,809,851 *nifH* sequences from 407 samples collected throughout the Atlantic Ocean (Figure 3.1) that contained 2655 operational taxonomic units (OTUs) when clustered at 96% sequence identity. Typically, rarefaction curves approached a plateau at a sequencing depth of 3000 reads (Figure 3.2), which corresponded on average to 84 ± 50 unique OTUs. This

indicates that we were able to capture the diazotrophic diversity in each sample to the extent that the commonly used *nifH* primers allow (Zehr et al. 2001). The saturation of unique OTUs at 3000 reads was also supported by the richness estimator Chao1, which indicated an average of 122 unique OTUs in the tropical surface ocean, a quantity which was reached in most samples at a sequencing depth of 3000 reads (Figure 3.2 and Figure 3.3B). For downstream community analysis, we chose to work with samples that were rarefied to 1500 reads, because this selected cutoff allowed the inclusion of samples from high latitudes with lower sequence counts and diversity than tropical samples, but which are historically from undersampled geographical regions with regards to diazotrophic diversity (Supplemental Table 12). The rarefaction cutoff did however lead to the exclusion of 29 samples; 25 originated from the GEOVIDE cruise transect (north of 40°N), and 4 originated from the Meteor 116 cruise transect.

Overall, *nifH* read numbers were much lower in the temperate ocean and the number of unique OTUs saturated much earlier on the rarefaction curve (on average at 20.4 OTUs; Figure 3.2) than in the tropics, already indicating that a less abundant and diverse diazotrophic community is present in those regions. Diversity analysis supported this finding; Shannon diversity and Chao1 estimator values were lowest north of 50°N (Figure 3.3), whereas the highest diversity was found in samples taken in the Bedford Basin in winter (> 300 OTUs). The Bedford Basin is a temperate oceanic inlet on the Canadian coast which experiences very different conditions to open ocean environments. It receives fresh water input from the Sackville River and rain water run-off, reaching an average salinity of 33

(Li 2014). Overall nutrient concentrations are high, although surface water fixed nitrogen concentrations become depleted in the summer months. The deep-water is low in O_2 concentrations, sometimes reaching suboxic or anoxic conditions in winter (Li 2014). These conditions possibly lead to an environment that is suitable for a wider range of diazotrophic organisms.

All diazotrophs depend on a sufficient supply of phosphate and iron, as well as an adequate carbon supply and possibly low O_2 concentrations (reviewed by Benavides et al. 2015 and Riemann et al. 2010). These requirements derive from the high iron content in the nitrogenase enzyme, the high energy requirements for the fixation of N_2 and the fact that the active site of the nitrogenase is extremely sensitive to oxidation (Fernandez et al. 2015; Riemann et al. 2010; Moisander et al. 2008; Berman-Frank et al. 2007; Kustka et al. 2003; Berman-Frank et al. 2001; Summons et al. 1999; Falkowski 1997). In 2014, the Bedford Basin phosphate concentrations exceeded fixed nitrogen concentrations throughout the year (N* was negative in 175 out of 196 samples taken in 2014; personal communication Richard Davis, CERC.OCEAN Dalhousie University) and seasonally low deep-water O_2 concentrations were observed in the winter months (O_2 concentrations below 90 µM in weeks 3-4 and 44 – 51, personal communication Richard Davis, CERC.OCEAN Dalhousie University). The dominant organisms in the diazotrophic community of the Bedford Basin were phylogenetically assigned to sub-cluster Ip (73.5% of sequences).

Throughout the entire dataset (Figure 3.4 and Supplemental Figure 5), all clusters, except sub-cluster Ic, displayed the highest relative abundances at

depths of 20 – 200 m. This included samples down to the bottom of the mixed layer, reaching the deep chlorophyll maximum where a supply of remineralized nutrients from below may sustain the microbial community including diazotrophs. Additionally, it has been proposed that heterotrophic diazotrophs may find a low O_2 concentration niche on sinking organic particles where respiration is high due to rapid degradation (Rieman et al. 2010).

High diversity was also observed throughout the tropical North Atlantic in the highly-oxygenated surface ocean (Meteor 116 cruise and Kn-204 surface samples from Cape Verde) and samples from the South Atlantic (Polarstern ANTXXVI-I), where surface nutrient concentrations were depleted (Figure 3.3). The South Atlantic Ocean and the tropical North Atlantic Ocean were inhabited by very different communities. The South Atlantic Ocean was dominated by *Candidatus* A. thalassa sequences, while *Trichodesmium spp.* was the dominant cyanobacterium in the tropical North Atlantic Ocean (Figure 3.4A). Diversity of the *nifH* gene decreased with depth and at latitudes above 40°N as the diazotrophic community changed as discussed below (Figure 3.3B and Figure 3.4).

3.4.2. Phylogenic affiliation of the retrieved **nifH** *OTUs*

The retrieved OTUs were classified under three major clusters (clusters I – III) previously described by Zehr et al. (2003), who found that cluster I includes cyanobacterial and most of the proteobacterial *nifH* sequences (sub-cluster Ic and sub-cluster Ip), cluster II contains archaeal and alternative *anfH* sequences and cluster III includes sequences from mainly anaerobic microorganisms. In

agreement with the finding of Farnelid et al. (2011) in a global ocean survey at 21 stations, the sub-cluster Ip contained the majority of recovered unique OTUs (53%; Figure 3.4). In agreement with recent investigations, sub-cluster Ip, cluster II and cluster III were dominant north of 40°N and in deep waters (Cheung et al. 2016; Langlois et al. 2015; Loescher et al. 2014; Moisander et al. 2014; Bird et al. 2013; Bonnet et al. 2013; Farnelid et al. 2011; Fernandez et al. 2011; Hewson et al. 2007; Bird et al. 2005; Metha et al. 2003). These studies hypothesized that non-cyanobacterial diazotrophs may dominate the community in more temperate, subsurface and in O_2 depleted water masses where growth conditions are not ideal for cyanobacterial species. However, we detected *nifH* genes related to *Trichodesmium* below the photic zone. Sinking of highly abundant cyanobacteria in the surface may skew the relative abundances of the actual thriving microbial communities at depth (Davis and McGillicuddy 2006). In order to determine whether *Trichodesmium* continues to fix N_2 at depth, analysis of *nifH* transcripts could be used to distinguish between the active and inactive diazotrophic community.

It has been suggested that the widely-used degenerate *nifH* primers (Zehr et al. 2001) are biased towards gamma-proteobacterial sequences and, as a result, the large sub-cluster Ip may be overrepresented (Turk-Kobo et al. 2014; Turk et al. 2011). Certainly, the nested *nifH* PCR utilizes degenerate primers, so primer bias is likely to occur and has also been shown to do so in specific cases (von Wintzingerode et al. 1997; Suzuki et al. 1996). However, we have attempted to minimize primer bias by performing PCRs reactions at two template dilutions in

the initial PCR, which is recommended in the standard Illumina MiSeq protocol for 16S rRNA gene amplification. The prevalence of cyanobacterial *nifH* sequences in the expected regions indicates that relative abundances roughly correlate with qPCR results from the same region (Moore et al., submitted, personal communication). Specific qPCR TaqMan assays have shown that Gamma A (AY896371), a gamma-proteobacterium, is widely distributed and actively expresses *nifH* throughout the oceans (Langlois et al. 2015). In our study, *nifH* sequences 100% identical to Gamma A (AY896371) were detected at 127 stations between 0 and 40°N (with a maximum of 276 reads per sample after rarefaction), the same region in which highest abundances were reported using clade-specific qPCR TaqMan assay (Langlois et al. 2015). The entire clade of Gamma A related sequences spread from 40°S to 40°N and from the surface to below 200 m, but never was found above 40°N (Figure 3.4B). Likewise, a recently isolated heterotrophic diazotroph from an oceanic inlet on the Canadian Atlantic coast (Bedford Basin) of sub-clade Ip (Chapter 5) was widely distributed throughout temperate oceanic regions, especially above 40°N, whereas relative abundances were lower in the tropical North Atlantic ocean as was also observed on the basis of its 16S rRNA gene sequence (Figure 3.4B). A TaqMan qPCR assay targeting this novel isolate correlated well with the recovery of its *nifH* phylotype from the large pool of sequences obtained in this study (Chapter 5). Hence, although some bias towards certain *nifH* sequences may be possible, it cannot account for the very high diversity of the sub-cluster Ip and the observed distribution patterns. Finally, earlier investigations into the diversity of marine *nifH* genes had shed doubt on the importance of non-cyanobacterial diazotrophs in

the ocean based on the recovery of *nifH* sequences contaminants found in PCR reagents and common laboratory environments (Turk et al. 2011). However, none of the contaminating sequences identified by Turk et al. (2011; AB198373, AB198377, AB198382, AB198384, AB198391, AY225107, EU916368 and EU916669) were recovered from our *nifH* sequence dataset.

Compared to Farnelid et al. (2011), we did find a large proportion of unique cluster II OTUs present (Figure 3.4). Farnelid et al. (2011) did not record OTUs in cluster II. At 96% identity, cluster II was comprised of 313 unique OTUs, making up a large portion in samples collected in the southern hemisphere and in the temperate ocean – areas that were not targeted by Farnelid et al. (2011; Figure 3.4 and Supplemental Figure 5). Cluster II contains sequences of archaeal origin or the alternative nitrogenase *anfH* (Zehr et al. 2003). There has been even less work on marine diazotrophic archaea or alternative nitrogenases than on proteobacteria, but recent evidence indicates that they are widely distributed and even though their role in the N_2 fixing community is not clear, it may be significant (McRose et al. 2017). Archaeal diazotrophs have been found in low nitrogen hydrothermal vent communities and sediments (Mehta et al. 2006; Dekas et al. 2005; Mehta et al. 2003), but have rarely been recovered from Sanger sequencing elsewhere in the ocean (Zehr et al. 2011). In this study, high relative abundances of cluster II sequences were found throughout the Polarstern ANTXXVI-I and GEOVIDE cruises (20 – 40°S and 40 – 60°N respectively; Figure 3.4C). Entire clades in each of *nifH* clusters II and III phylogenetic trees were almost exclusively found in the South Atlantic (Figure 3.4C and D). The cluster II

sequences were ca. 86% similar to *nifH* sequences found in soils of the Antarctic Dry Valleys (Niederberger et al. 2012) and the cluster III sequences were ca. 93% similar to sequences previously found in the Central Arctic Ocean (Fernández-Méndez et al. 2016). Diversity in the GEOVIDE samples was low, which suggests that the high relative abundance of cluster II sequences in those samples points to the presence of only a few diazotrophs of cluster II in that region.

Cluster III OTUs were found throughout all samples, but only dominated the community in 10 samples, which were located on the Nova Scotian coast (5 samples, AZMP HUD2014004 and HUD2014030 cruises), the most northern GEOVIDE samples near Greenland (3 samples) and near Cape Verde (2 samples; Supplemental Figure 5). We did not find the high prevalence of cluster III sequences in high latitudes of the North Atlantic Ocean observed by Farnelid et al. (2011). Although cluster III made up a larger portion of the diazotrophic community in the temperate North Atlantic and contained many OTUs solely found there, sub-cluster Ip sequences were detected most frequently north of 40°N (Figure 3.4D). This was also the case for samples collected in the Arctic Ocean (Fernández-Méndez et al. 2016). Many OTUs of sub-cluster Ip were almost exclusively associated with the temperate ocean and clustered far from reference genomes, which underlines an important research gap associated with diazotrophs in the temperate North Atlantic Ocean (Figure 3.4B).

Cyanobacterial OTUs were found in 356 of the 407 samples, and dominant OTUs in 262 samples were of cyanobacterial origin (Figure 3.5). This is because most

samples were collected in the tropical Atlantic Ocean between 0 and 20°N and OTUs belonging to *Trichodesmium* and *Candidatus* A. thalassa were dominant in this region (Table 3.1: Sample numbers for oceanic regions and depths.; Figure 3.7). Figure 3.5 also shows that the majority of OTUs were detected in less than 50 samples, but ranged from only 1 to over 10,000 detected reads in the rarefied dataset. However, samples in spatial studies of OTU occurrence and distribution represent a snapshot over a finite time span. The high relative abundance of specific OTUs could be connected to blooming events, the absence of other sequences or through the induction by specific environmental events. Although the spatial coverage achieved in this study was useful in assessing the geographical range of *nifH* OTUs common to a broad range of samples, we cannot assess the absolute abundance of their *nifH* phylotypes in individual samples. This would require the design of phylotypes specific qPCR assays. In addition, a spatial study cannot answer the question of whether specific OTUs are always abundant in an environment and whether rare OTUs may at times dominate throughout large oceanic regions. In this respect, time-series studies of marine microbial communities have shown that communities vary inter-annually, seasonally or even weekly in response to environmental changes (Giovanni et al. 2012; Gilbert et al. 2011; Chapter 6). Only time series studies can draw conclusions about the changes and dominance of diazotrophic species with changing environmental conditions.

3.4.3. Diazotrophic community throughout the Atlantic Ocean

Bray-Curtis Similarity plots of our sample pool revealed two major clusters of diazotrophic communities, with a clear separation of communities north of 40°N (Figure 3.6A). This community shift was also seen in the relative abundances of *nifH* clusters (Supplemental Figure 5). 40°N marks the border between samples influenced by the North Atlantic gyre or by water masses derived from Arctic currents, each with their own nutrient content and their own selection pressures on diazotrophic organisms. A few samples from the AZMP HUD2014004 and HUD2014030 clustered with the more southern samples, which originated from the southern part on the Scotian Shelf. This part is connected to the North Atlantic gyre via the Gulf Stream which may explain the similarity of community to the southern samples.

To investigate the difference between clusters, we extracted the single most commonly found OTU from each sample and, when possible, used BLAST to find the nearest sequenced relative (Altschul et al. 1998). Besides the most commonly found *Trichodesmium* and *Candidatus* A. thalassa (Figure 3.7), the next most often recorded reference genome was *Magnetococcus marinus* also in the tropical North Atlantic. However, the sequence identity was only 78 – 80%, which indicated a very distant relationship. Therefore, taxonomic assignment cannot be securely determined. The most abundant OTUs from samples above 40°N were much more variable, showing overall lower identity with sequenced diazotrophic genomes and only rarely were part of the cyanobacterial sub-cluster (Figure 3.7 and Supplemental Figure 5), supporting Non-linear Multi-Dimensional

(NMDS) community plot results (Figure 3.6). Diazotrophic OTUs identified north of 40°N were all part of sub-cluster Ip: *Agarivorans gilvus* (99% identity, isolated from seagrass on the Chinese coast (Du et al. 2011)), *Klebsiella oxytoca* (96% identity, a human pathogen, but also associated with agricultural plants (Adachi et al. 2002; Podschun et al. 1998)) and *Marinobacterium litorale* (99% identity, isolated from the coastal Yellow Sea (Kim et al. 2007)). The isolated diazotroph (characterized in Chapter 5) was found to be the dominant OTU in five samples of the GEOVIDE cruise (> 42°N, 20 – 200 m), but was also occasionally found in the tropical ocean.

Further analysis of the samples collected south of 40°N indicated some separation according to depth and latitude either north of 20°N or south of 20°S (Figure 3.6B). Cyanobacterial dominated samples clustered tightly together (Figure 3.6B and Figure 3.7). In comparison, sub-cluster Ip sequences were more abundant in samples further north or at depth, whereas cluster II and III sequences were present in samples further south (Figure 3.4 and Supplemental Figure 5). The most southern samples were collected in waters influenced by the Brazil current, which forms the western border of the Southern gyre. *Candidatus* A. thalassa dominated the cyanobacterial community in those samples, and the large proportion of clusters II and III distinguished these samples from the tropical North Atlantic (Figure 3.7). One of the cluster III OTUs was the dominant OTU in 6 of 12 samples. Its closest reference genome was found to be *Desulfovibrio alkalitolerans* (80% sequence identity). This OTU has not been previously

deposited in the NCBI database and was only found in the South Atlantic cruise transect along with several other unique OTUs from this region (Figure 3.4C).

From this overview, we find supportive evidence for previous findings: Cyanobacterial species are bound to the sunlit ocean and to areas of warm to moderate temperatures, as we did not record many cyanobacterial sequences at latitudes higher than 40°N in the northern cruise transects (Figure 3.4 and Supplemental Figure 5), which marks the border of the North Atlantic gyre and with it, dramatically changing environmental conditions. Samples north of the gyre were markedly less diverse and contained completely different communities than those within the gyre (Figure 3.6 and Supplemental Figure 5). Samples below the surface were more likely to be dominated by proteobacteria than surface samples, though probable transport of surface organisms through sinking material or mixing could confound community structure analysis (Figure 3.6 and Supplemental Figure 5). The cruise section in the South Atlantic contained a specific community composed of *Candidatus* A. thalassa, and poorly characterized OTUs associated with clusters II and III. From this, we concluded that non-cyanobacterial diazotrophs are subjected to different environmental constraints than cyanobacterial diazotrophs. While Bentzon-Tilia et al. (2014) showed that three species recently isolated from the brackish Baltic Sea each showed specific requirements for nutrient and O_2 concentrations (2014), which points to the potential diversity of metabolism that is present in the non-cyanobacterial diazotrophic community, there is still a dearth of information on the growth condition, function and adaptations of non-cyanobacterial diazotrophs.

It is clear that further studies are needed to shed light on the factors that shape the very diverse marine diazotrophic communities.

A purely DNA-derived study is limited in that it only indicates the presence of particular organisms. However, the presence of these organisms does not allow conclusions about whether they are actively fixing N_2. Several studies support the assumption that cyanobacteria in the tropical Atlantic Ocean are fixing N_2 (reviewed by Luo et al. 2012), but little is understood about the N_2 fixation rates, regulation and environmental requirements of non-cyanobacterial diazotrophs as well as their global impact (Benavides et al. 2015; Luo et al. 2012). There is a pressing need to establish whether the diverse and proportionally very abundant group of non-cyanobacterial diazotrophs contributes significantly to global N_2 fixation. This may help reconcile biological N_2 fixation rate calculations with geochemical estimates (Codispoti 2007; Altabet 2006; Mahaffey et al. 2005).

3.4.4. *Symbiotic relationships of* Candidatus *Atelocyanobacterium thalassa*

The unicellular cyanobacterial diazotroph *Candidatus* A. thalassa has drawn attention because it lacks essential genes to perform all required steps during photosynthesis and carbon fixation and was therefore assumed to be a symbiont (Zehr et al. 2008). *Candidatus* A. thalassa has been found in association with a haptophyte host, specifically *Braarudosphaera bigelowii* (Thompson et al. 2012). Additionally, Thompson et al. (2014) proposed that clade 2 of the three clades of *Candidatus* A. thalassa associates with a different host than clades 1 and 3. We

used SparCC to search for a correlation between *Candidatus* A. thalassa OTUs and chloroplast derived 16S rRNA gene sequences (Table 3.2, Figure 3.8). Our sequence-derived results are in accordance those of Thompson et al. (2014), who established the differences in hosts using fluorescent activated cell sorting (FACS) and PCR. In our study, clades 1 and 3 were most strongly correlated with Haptophyceae OTU 5 (Hapt NROTU5), whereas clade 2 only correlated with Haptophyceae OTU 32 (Hapt NROTU32). Both OTUs have 99% sequence identity to *Braarudosphaera bigelowii* and cluster closely together in phylogenetic analysis (Supplemental Figure 9). *Candidatus* A. thalassa clade 1 additionally correlated with OTU 137837 of the Stramenopiles (Stram 137837), which the symbiotic *Richelia* is also associated with (Villareal 1990), supporting a potential for possible symbiosis. However, the correlation may be based on the wide distribution of Stramenopiles in the tropical Atlantic Ocean rather than a symbiotic relationship (Massana et al. 2004).

3.5. Conclusion

Building on the increasing application of high-throughput sequencing to exploring diazotrophic diversity, we conducted a thorough study of the *nifH* gene in the Atlantic Ocean using Illumina pair-end sequencing. We observed the previously established dominance of cyanobacterial diazotrophs in the tropical surface ocean (Luo et al. 2012) and found evidence supporting the presence of a non-cyanobacterial dominated community with distinctive distribution patterns outside of these areas. The non-cyanobacterial clusters were mainly comprised of unique

sequences, and we propose that research into these organisms is needed to obtain a global picture of the marine diazotrophic community. Additionally, the correlation of our *nifH* dataset with chloroplast 16S rRNA genes from the same samples showed that clade 2 of *Candidatus* A. thalassa lives in symbiosis with a different host than clade 1 and 3, which supports the findings of Thompson et al. (2014). Although more work is required to confirm the *Candidatus* A. thalassa host, our study demonstrates the potential application of high-throughput sequencing to identify possible symbiotic relationships between diazotrophs and host organisms.

3.6. Acknowledgements

Jenni-Marie Ratten's work was supported by the NSERC CREATE Transatlantic Ocean System Science and Technology (TOSST) grant, by the Deutscher Akademischer Austausch Dienst (DAAD), the Stiftunf für Kanadastudiesn (SKS), and Julie LaRoche's NSERC Discovery research grant.

3.7. References

Adachi, K., Nakatani, M. and Mochida, H., 2002. Isolation of an endophytic diazotroph, Klebsiella oxytoca, from sweet potato stems in Japan. Soil science and plant nutrition, 48(6), pp.889-895.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

Altabet, M.A., 2006. Constraints on oceanic N balance/imbalance from sedimentary 15 N records. Biogeosciences Discussions, 3(4), pp.1121-1155.

Benavides, M., Luo, Y.W., Doney, S.C., Anderson, L.A., Bode, A., Bonnet, S., Boström, K.H., Böttjer, D., Capone, D.G., Carpenter, E.J. and Chen, Y.L., 2012. Database of diazotrophs in global ocean: abundances, biomass and nitrogen fixation rates.

Benavides, M. and Voss, M., 2015. Five decades of N_2 fixation research in the North Atlantic Ocean. Frontiers in Marine Science, 2, p.40.

Bentzon-Tilia, M., Farnelid, H., Jürgens, K. and Riemann, L., 2014. Cultivation and isolation of N_2-fixing bacteria from suboxic waters in the Baltic Sea. FEMS microbiology ecology, 88(2), pp.358-371.

Bentzon-Tilia, M., Traving, S.J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J.L., Markager, S. and Riemann, L., 2015. Significant N2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. The ISME journal, 9(2), pp.273-285.

Berman-Frank, I., Lundgren, P. and Falkowski, P., 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Research in microbiology, 154(3), pp.157-164.

Berman-Frank, I., Quigg, A., Finkel, Z.V., Irwin, A.J. and Haramaty, L., 2007. Nitrogen-fixation strategies and Fe requirements in cyanobacteria. Limnology and Oceanography, 52(5), pp.2260-2269.

Bird, C., Martinez, J.M., O'Donnell, A.G. and Wyman, M., 2005. Spatial distribution and transcriptional activity of an uncultured clade of planktonic diazotrophic γ-Proteobacteria in the Arabian Sea. Applied and environmental microbiology, 71(4), pp.2079-2085.

Bird, C. and Wyman, M., 2013. Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea. FEMS microbiology ecology, 84(1), pp.189-200.

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K.A., Moutin, T., Hamersley, R.M., Grosso, O., Zehr, J.P. and Capone, D.G., 2013. Aphotic N 2 fixation in the eastern tropical South Pacific Ocean. PloS one, 8(12), p.e81265.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I. and Huttley, G.A., 2010a. QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7(5), pp.335-336.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R., 2010b. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics, 26(2), pp.266-267.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), pp.540-552.

Cheung, S., Xia, X., Guo, C. and Liu, H., 2016. Diazotroph community structure in the deep oxygen minimum zone of the Costa Rica Dome. Journal of plankton research, 38(2), pp.380-391.

Clarke, K.R. and Gorley, R.N., 2006. PRIMER V6: user manual-tutorial. Plymouth Marine Laboratory.

Codispoti, L.A., 2007. An oceanic fixed nitrogen sink exceeding 400 Tg N a? 1 vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences, 4(2), pp.233-253.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W. and Yoshinari, T., 2001. The oceanic fixed nitrogen and nitrous oxide budgets: moving targets as we enter the anthropocene?. Scientia Marina, 65(S2), pp.85-105.

Comeau, A.M., Douglas, G.M. and Langille, M.G., 2017. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. mSystems, 2(1), pp.e00127-16.

Davis, C.S. and McGillicuddy, D.J., 2006. Transatlantic abundance of the N2-fixing colonial cyanobacterium *Trichodesmium*. Science, 312(5779), pp.1517-1520.

Decelle, J., Romac, S., Stern, R.F., Bendif, E.M., Zingone, A., Audic, S., Guiry, M.D., Guillou, L., Tessier, D., Le Gall, F. and Gourvil, P., 2015. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. Molecular ecology resources, 15(6), pp.1435-1445.

Dekas, A.E., Poretsky, R.S. and Orphan, V.J., 2009. Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. Science, 326(5951), pp.422-426.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14(6), pp.927-930.

Du, Z.J., Lv, G.Q., Rooney, A.P., Miao, T.T., Xu, Q.Q. and Chen, G.J., 2011. Agarivorans gilvus sp. nov. isolated from seaweed. International journal of systematic and evolutionary microbiology, 61(3), pp.493-496.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), pp.1792-1797.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27(16), pp.2194-2200.

Falkowski, P.G., 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. Nature, 387(6630), p.272.

Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å. and Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PloS one, 6(4), p.e19223.

Farnelid, H., Bentzon-Tilia, M., Andersson, A.F., Bertilsson, S., Jost, G., Labrenz, M., Jürgens, K. and Riemann, L., 2013. Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. The ISME journal, 7(7), pp.1413-1423.

Fernandez, C., Farías, L. and Ulloa, O., 2011. Nitrogen fixation in denitrified marine waters. PloS one, 6(6), p.e20539.

Fernandez, C., González, M.L., Muñoz, C., Molina, V. and Farias, L., 2015. Temporal and spatial variability of biological nitrogen fixation off the upwelling system of central Chile (35–38.5° S). Journal of Geophysical Research: Oceans, 120(5), pp.3330-3349.

Fernández-Méndez, M., Turk-Kubo, K.A., Buttigieg, P.L., Rapp, J.Z., Krumpen, T., Zehr, J.P. and Boetius, A., 2016. Diazotroph Diversity in the Sea Ice, Melt Ponds, and Surface Waters of the Eurasian Basin of the Central Arctic Ocean. Frontiers in Microbiology, 7.

Friedman, J. and Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol, 8(9), p.e1002687.

Gamble, M.D., Bagwell, C.E., LaRocque, J., Bergholz, P.W. and Lovell, C.R., 2010. Seasonal variability of diazotroph assemblages associated with the rhizosphere of the salt marsh cordgrass, *Spartina alterniflora*. Microbial ecology, 59(2), pp.253-265.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I. and Somerfield, P., 2012. Defining seasonal marine microbial community dynamics. The ISME journal, 6(2), pp.298-308.

Giovannoni, S.J. and Vergin, K.L., 2012. Seasonality in ocean microbial communities. Science, 335(6069), pp.671-676.

Gruber, N., 2008. The marine nitrogen cycle: overview and challenges. Nitrogen in the marine environment, 2, pp.1-50.

Hamersley, M.R., Lavik, G., Woebken, D., Rattray, J.E., Lam, P., Hopmans, E.C., Damsté, J.S.S., Krüger, S., Graco, M., Gutiérrez, D. and Kuypers, M.M., 2007. Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. Limnology and Oceanography, 52(3), pp.923-933.

Hewson, I., Moisander, P.H., Achilles, K.M., Carlson, C.A., Jenkins, B.D., Mondragon, E.A., Morrison, A.E. and Zehr, J.P., 2007. Characteristics of diazotrophs in surface to abyssopelagic waters of the Sargasso Sea. Aquatic microbial ecology, 46(1), pp.15-30.

Hoeft McCann, S., Boren, A., Hernandez-Maldonado, J., Stoneburner, B., Saltikov, C.W., Stolz, J.F. and Oremland, R.S., 2016. Arsenite as an Electron Donor for Anoxygenic Photosynthesis: Description of Three Strains of *Ectothiorhodospira* from Mono Lake, California and Big Soda Lake, Nevada. Life, 7(1), p.1.

Ingall, E. and Jahnke, R., 1994. Evidence for enhanced phosphorus regeneration from marine sediments overlain by oxygen depleted waters. Geochimica et Cosmochimica Acta, 58(11), pp.2571-2575.

Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research, 30(14), pp.3059-3066.

Kim, H., Choo, Y.J., Song, J., Lee, J.S., Lee, K.C. and Cho, J.C., 2007. Marinobacterium litorale sp. nov. in the order Oceanospirillales. International journal of systematic and evolutionary microbiology, 57(7), pp.1659-1662.

Kong, L., Jing, H., Kataoka, T., Sun, J. and Liu, H., 2011. Phylogenetic diversity and spatio-temporal distribution of nitrogenase genes (nifH) in the northern South China Sea. Aquatic Microbial Ecology, 65(1), pp.15-27.

Kopylova, E., Noé, L. and Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics, 28(24), pp.3211-3217.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution, p.msw054.

Kustka, A.B., Sanudo-Wilhelmy, S.A., Carpenter, E.J., Capone, D., Burns, J. and Sunda, W.G., 2003. Iron requirements for dinitrogen-and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations. Limnology and Oceanography, 48(5), pp.1869-1884.

Langlois, R., Großkopf, T., Mills, M., Takeda, S. and LaRoche, J., 2015. Widespread distribution and expression of gamma A (UMB), an uncultured, diazotrophic, γ-proteobacterial nifH phylotype. PloS one, 10(6), p.e0128912.

Letunic, I. and Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, 44(W1), pp.W242-W245.

Li, W. K., 2014. The state of phytoplankton and bacterioplankton at the Compass Buoy Station: Bedford Basin Monitoring Program 1992-2013. Canadian Technical Report of Hydrography and Ocean Sciences 304.

Loescher, C.R., Großkopf, T., Desai, F.D., Gill, D., Schunck, H., Croot, P.L., Schlosser, C., Neulinger, S.C., Pinnow, N., Lavik, G. and Kuypers, M.M., 2014. Facets of diazotrophy in the oxygen minimum zone waters off Peru. The ISME journal, 8(11), pp.2180-2192.

Mahaffey, C., Michaels, A.F. and Capone, D.G., 2005. The conundrum of marine N_2 fixation. American Journal of Science, 305(6-8), pp.546-595.

Martinez-Perez, C., Mohr, W., Löscher, C.R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B.M., Lavik, G., Schmitz, R.A. and LaRoche, J., 2016. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. Nature Microbiology, 1, p.16163.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K. and Pedrós-Alió, C., 2004. Phylogenetic and ecological analysis of

novel marine stramenopiles. Applied and environmental microbiology, 70(6), pp.3528-3534.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. and Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal, 6(3), pp.610-618.

McRose, D.L., Zhang, X., Kraepiel, A.M. and Morel, F.M., 2017. Diversity and activity of alternative nitrogenases in sequenced genomes and coastal environments. Frontiers in Microbiology, 8, p.267.

Mehta, M.P. and Baross, J.A., 2006. Nitrogen fixation at 92 C by a hydrothermal vent archaeon. Science, 314(5806), pp.1783-1786.

Mehta, M.P., Butterfield, D.A. and Baross, J.A., 2003. Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge. Applied and Environmental Microbiology, 69(2), pp.960-970.

Mercier, C., Boyer, F., Bonin, A. and Coissac, E., 2013, November. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. In Programs and Abstracts of the SeqBio 2013 workshop. Abstract (pp. 27-29).

Moisander, P.H., Beinart, R.A., Voss, M. and Zehr, J.P., 2008. Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. The ISME journal, 2(9), pp.954-967.

Moisander, P.H., Serros, T., Paerl, R.W., Beinart, R.A. and Zehr, J.P., 2014. Gammaproteobacterial diazotrophs and nifH gene expression in surface waters of the South Pacific Ocean. The ISME journal, 8(10), pp.1962-1973.

Niederberger, T.D., Sohm, J.A., Tirindelli, J., Gunderson, T., Capone, D.G., Carpenter, E.J. and Cary, S.C., 2012. Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys. FEMS microbiology ecology, 82(2), pp.376-390.

Podschun, R. and Ullmann, U., 1998. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clinical microbiology reviews, 11(4), pp.589-603.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing.

Ratten, J.M., LaRoche, J., Desai, D.K., Shelley, R.U., Landing, W.M., Boyle, E., Cutter, G.A. and Langlois, R.J., 2015. Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. Deep Sea Research Part II: Topical Studies in Oceanography, 116, pp.332-341.

Riemann, L., Farnelid, H. and Steward, G.F., 2010. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. Aquatic Microbial Ecology, 61(3), pp.235-247.

Schlitzer, R., 2015. Ocean Data View. odv. awi. de.

Severin, I., Bentzon-Tilia, M., Moisander, P.H. and Riemann, L., 2015. Nitrogenase expression in estuarine bacterioplankton influenced by organic carbon and availability of oxygen. FEMS microbiology letters, 362(14), p.fnv105.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), pp.1312-1313.

Summons, R.E., Jahnke, L.L., Hope, J.M. and Logan, G.A., 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature, 400(6744), pp.554-557.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. and Cornejo-Castillo, F.M., 2015. Structure and function of the global ocean microbiome. Science, 348(6237), p.1261359.

Suyama, M., Torrents, D. and Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research, 34(suppl 2), pp.W609-W612.

Suzuki, M.T. and Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and environmental microbiology, 62(2), pp.625-630.

Tamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution, 24(8), pp.1596-1599.

Thompson, A., Carter, B.J., Turk-Kubo, K., Malfatti, F., Azam, F. and Zehr, J.P., 2014. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. Environmental microbiology, 16(10), pp.3238-3249.

Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaulot, D., Kuypers, M.M. and Zehr, J.P., 2012. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. Science, 337(6101), pp.1546-1550.

Turk, K.A., Rees, A.P., Zehr, J.P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E.M.S. and Gilbert, J., 2011. Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. The ISME journal, 5(7), pp.1201-1212.

Turk-Kubo, K.A., Karamchandani, M., Capone, D.G. and Zehr, J.P., 2014. The paradox of marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. Environmental microbiology, 16(10), pp.3095-3114.

Villareal, T.A., 1990. Laboratory culture and preliminary characterization of the nitrogen-fixing Rhizosolenia-Richelia symbiosis. Marine Ecology, 11(2), pp.117-132.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology, 73(16), pp.5261-5267.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R. and Ley, R.E., 2012. Impact of

training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME journal, 6(1), pp.94-103.

Wintzingerode, F.V., Göbel, U.B. and Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS microbiology reviews, 21(3), pp.213-229.

Xiao, P., Jiang, Y., Liu, Y., Tan, W., Li, W. and Li, R., 2015. Re-evaluation of the diversity and distribution of diazotrophs in the South China Sea by pyrosequencing the nifH gene. Marine and Freshwater Research, 66(8), pp.681-691.

Yamada, K.D., Tomii, K. and Katoh, K., 2016. Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. Bioinformatics, p.btw412.

Zehr, J.P., 2011. Nitrogen fixation by marine cyanobacteria. Trends in microbiology, 19(4), pp.162-173.

Zehr, J.P., Bench, S.R., Carter, B.J., Hewson, I., Niazi, F., Shi, T., Tripp, H.J. and Affourtit, J.P., 2008. Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. Science, 322(5904), pp.1110-1112.

Zehr, J.P., Jenkins, B.D., Short, S.M. and Steward, G.F., 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental microbiology, 5(7), pp.539-554.

Zehr, J.P. and Turner, P.J., 2001. Nitrogen fixation: nitrogenase genes and gene expression. Methods in microbiology, 30, pp.271-286.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), pp.614-620.

CHAPTER 4: INFERRING THE METABOLIC DIVERSITY OF MARINE NON-CYANOBACTERIAL DIAZOTROPHS

Jenni-Marie Ratten, Dhwani Desai, Julie LaRoche

Contribution of authors:

Jenni-Marie Ratten:	Collection of data, data analysis, drafting of the manuscript
Dhwani Desai	Support with statistical analysis
Julie LaRoche	Planning and discussion of the manuscript

4.0. Abstract

Recent findings suggest that the diazotrophic community outside of the cyanobacteria-dominated tropical surface oceans may harbour completely different species compositions. Non-cyanobacterial diazotrophs were found to dominate the diazotrophic community in many marine environments, including the aphotic zone, temperate oceans, and O_2-depleted waters, suggesting that these diazotrophs may contribute significantly to global marine N_2 fixation. However, we know very little about how these organisms regulate their metabolism, including N_2 fixation and intracellular O_2 concentrations, and very few

non-cyanobacterial marine diazotrophs have been cultivated, thus limiting the ability to study the physiological ecology of the dominant marine groups in controlled experiments. The increasing number of diazotrophic genomes from diverse environments can provide some insight into their metabolic potential and that of closely related marine diazotrophs. This review summarizes our knowledge of the distribution and abundance of non-cyanobacterial diazotrophs and infers their additional metabolic potential from the genomes of 132 diazotrophs closely aligned with marine diazotrophic *nifH* phylotypes. Our analysis demonstrates the extensive potential metabolic diversity of diazotrophs. Inferring from their gene complements, metabolic potential differed significantly between taxonomic groups. In particular, firmicute and cyanobacterial metabolisms differed from the rest of the diazotrophs, mainly driven by genes of obligate anaerobic respiration and genes involved in photosynthesis, respectively. Metabolic diversity was highest in alpha- , beta- and gamma-Proteobacteria and lowest among Archaea, delta- and epsilon-Proteobacteria, and Firmicutes. Alongside a set of core metabolic pathways present in all organisms, various groups of diazotrophs could utilize alternative organic carbon sources, degrade aromatic compounds and use phosphonate. The potential to carry out anaerobic respiration using a variety of electron donors and acceptors was found in most genomes. Various taxonomic groups and in particular *Trichodesmium spp.* and *Actinobacteria*, harbored genes for elaborate iron-acquisition systems, including siderophore assembly as well as transport and processing of heme and hemin. Several diazotrophs possess genes for nitrate and nitrite assimilation in addition to the full *nif* operon. More surprisingly, four of

the 132 diazotrophs (alpha- and gamma-Proteobacteria) display the potential to respire fixed nitrogen anaerobically via the denitrification pathway, suggesting that feedback loops between gain and loss pathways may operate within a single microbial species. Further, 59 non-cyanobacterial reference genomes showed the potential for autotrophy, splitting the group of non-cyanobacterial diazotrophs into the subgroups of heterotrophs and autotrophs. Assuming that we can extrapolate from cultivated strains to marine strains based on a classification of *nifH* gene identity, the diverse metabolic potential of non-cyanobacterial diazotrophs can possibly explain the widespread distribution of these organisms throughout the global ocean and supports the idea that their role in the microbial community far exceeds N_2 fixation.

4.1. Introduction

The global oceanic inventory of dissolved inorganic nitrogen (DIN) is controlled by microbially-mediated loss and gain pathways. Biological N_2 fixation, carried out exclusively by diazotrophs, is the only natural pathway aside from atmospheric deposition and lightning that can bring new DIN into the open ocean, and counteracts the loss pathways of denitrification and anammox (Gruber 2008; Codispoti 2007; Hamersley et al. 2007; Codispoti et al. 2001; Ingall et al. 1994). Diazotrophs are a phylogenetically-diverse collection of bacteria and archaea that have the ability to reduce atmospheric N_2 gas to ammonia. Their nitrogenase enzyme is responsible for N_2 fixation and operates with two subunits, which are encoded by the *nifD/H/K* genes on the *nif* operon.

Chronic depletion of DIN in surface waters of permanently stratified oligotrophic subtropical gyres and seasonal depletion in several other temperate regions of the oceans led to nitrogen limitation of primary productivity in large oceanic areas, opening a niche for photosynthetic cyanobacterial diazotrophs (Moore et al. 2013). Photosynthetic diazotrophs, especially the microscopically identifiable *Trichodesmium spp.* and diatom-associated *Richelia,* have been studied extensively and are well-recognized contributors to marine N_2 fixation (Carpenter et al. 1999; Zehr et al. 1998; Capone et al. 1997).

However, since the 1990's, PCR amplification and Sanger sequencing of *nifH* genes has uncovered a much broader phylogenetic diversity of diazotrophs in the marine environment (Zehr 2011; Langlois et al. 2008; Needoba et al. 2007; Langlois et al. 2005; Karl et al. 2002; Zehr et al. 1998). More recently, high-throughput tag sequencing of the *nifH* gene has provided an even more detailed picture of the diversity, community structure, and distribution of marine diazotrophs, with representatives in *nifH* clusters I – III across the world oceans (Cheung et al. 2016; Bentzon-Tilia et al. 2015; Severin et al. 2015; Farnelid et al. 2013; Farnelid et al. 2011; Zehr et al. 2003a). These clusters include a wide taxonomic range of non-cyanobacterial diazotrophs, although most are currently solely classified according to their *nifH* gene sequence and thus we are without cultivated marine isolates representative of these novel *nifH* clades (Farnelid et al. 2011a; Riemann et al. 2010; Zehr et al. 2003a). Here, we summarize the peer-reviewed literature related to the diversity of *nifH* phylotypes, with a focus on the diversity of the non-cyanobacterial phylotypes.

Due to the lack of cultivated marine non-cyanobacterial diazotrophs, our current view of marine diazotrophs is founded on our knowledge of marine photosynthetic diazotrophs that thrive in tropical and subtropical regions depleted of DIN. While photosynthetic cyanobacterial diazotrophs are restricted to the euphotic zone, marine non-cyanobacterial diazotrophs are widely distributed in the world's oceans. Although they are also found in surface waters, their distribution range extends throughout the aphotic, temperate waters in both oxic and O_2-depleted oceanic waters indicating distinctly different metabolic adaptations to the environment (Langlois et al. 2015; Ratten et al. 2015; Loescher et al. 2014; Diez et al. 2012; Jayakumar et al. 2012; Farnelid et al. 2011b; Fernandez et al. 2011; Jungblut et al. 2010; Dekas et al. 2009; Rees et al. 2009; Stal 2009; Moisander et al. 2008; Holl et al. 2007; Needoba et al. 2007).

The phylogenetic diversity of the *nifH* genes recovered from the marine environment combined with the wide distribution throughout other environments (e.g. benthic environment, microbial mats, sea grass, coral reefs, tropical forest, termite guts and crop plants) suggest that the metabolic potential and therefore the ecological guild of these groups is likely very diverse (Tu et al. 2016; Cook et al. 2015; Andersson et al. 2014; Bertics et al. 2013; Vitousek et al. 2013; Desai et al. 2012; Lema et al. 2012). Hence, the contribution of N_2 fixation by non-cyanobacterial diazotrophs may have been overlooked, which, together with a systematic methodological underestimation of N_2 fixation measurements (Großkopf et al. 2012a), could explain the discrepancies in global N_2 fixation estimates within the marine nitrogen cycle (Codispoti 2007; Altabet 2006;

Mahaffey et al. 2005). Keeping in mind what is currently known of the distribution and phylogenetic affiliation of *nifH* phylotypes in the ocean, we analysed the genomes of closely related diazotrophs to provide an overview of the potential cellular metabolic pathways in addition to diazotrophy that may also operate in non-cyanobacterial marine diazotrophs.

4.2. Questions and challenges

When conducting research on diazotrophic organisms and interpreting current research studies, there are a few challenges to be addressed. The abundance of non-cyanobacterial *nifH* sequences, as measured with qPCR is generally three orders of magnitude lower (10^1 – 10^3 cells L^{-1}, rarely exceeding 10^5 copies L^{-1}) than cyanobacterial diazotrophs, making it difficult to detect non-cyanobacterial diazotrophs in clone libraries among cyanobacterial species (Langlois et al. 2015; Moisander et al. 2014; Moisander et al. 2010; Church et al. 2008; Langlois et al. 2008; Zehr et al. 2007; Church et al. 2005). The greater sequencing depth enabled by high-throughput sequencing of the *nifH* gene returns thousands of sequences per sample, facilitating the detection of non-cyanobacterial diazotrophs in almost all oceanic biomes (Farnelid et al. 2011). Estimating cell abundances is further confounded by the relationship between *nifH* copy number and actual cell numbers. It has been shown for *Trichodesmium spp.* that *nifH* counts determined by qPCR were up to ten times higher than microscopic counts due to polyploidy (Sargent et al. 2016). The number of *nifH* gene copies per cell for any given diazotroph cannot be determined until the genome has been

sequenced and the numbers of genomes per cell have been investigated, a caveat which must be considered by any study using nucleic acid-based quantification of *nifH*. If the 16S rRNA gene is known for a specific diazotroph, direct cell counts are achievable via Fluorescent In Situ Hybridisation (FISH) assays, as was recently demonstrated for the symbiotic *Candidatus* Atelocyanobacterium thalassa clade (Martinez-Perez et al. 2016). However the majority of non-cyanobacterial diazotrophs are identified solely by a *nifH* gene sequence for which no matching 16S rRNA gene sequence yet exists.

PCR bias, potentially increased through dual amplification rounds of the nested *nifH* PCR, may skew results towards an overestimation of gamma-proteobacterial diazotrophs (Turk-Kobo et al. 2014; Hewson et al. 2007; Turk et al 2011; von Wintzingerode et al. 1997; Suzuki et al. 1996); and the sinking of cells from the surface into the ocean's interior may further confound estimates of diazotroph abundance and distribution.

Metagenomics, which avoids the biases inherent in gene-specific identification, has been a successful approach for abundant marine microbial species; however, even the most abundant diazotrophs are orders of magnitude lower than the most common marine microorganisms, resulting in an extremely low occurrence in metagenome sequence data. Thus, standard metagenomics remain an inefficient technique unless the fraction of diazotrophs is enriched through other techniques such as cell-sorting flow cytometry (Hilton et al. 2014; Rahav et al. 2013b; Johnston et al. 2005).

We present a discussion of the recent scientific literature to present the evidence supporting the wide distribution of non-cyanobacterial diazotrophs throughout all oceanic environments. We also investigate the diverse metabolic potential of non-cyanobacterial diazotrophic supporting their adaptation mechanisms to a variety of environments, and show that their metabolic role far exceeds that of simply adding fixed nitrogen species to DIN-depleted oceanic regions.

4.3. Detection of non-cyanobacterial diazotrophs in the global oceans

4.3.1. Open Ocean

Until recently, the focal point of marine N_2 fixation research has been the surface waters of the tropical Atlantic and Pacific Oceans, which are dominated by cyanobacterial *nifH* phylotypes (Benavides et al. 2015; Luo et al. 2012). However, non-cyanobacterial members of the diazotrophic communities are still present in these waters. For example, although no representative strains have been isolated to date, a specific gamma-proteobacterial *nifH* gene (Gamma A; AY896371), first retrieved from the Arabian Sea (Bird et al. 2005), has been widely detected by a clade-specific TaqMan assay in more than 1000 DNA and cDNA samples from the tropical and subtropical waters of the Atlantic and Pacific Oceans, with numbers reaching up to 10^5 *nifH* copies L^{-1} (Langlois et al. 2015). More generally, pyrosequencing of *nifH* gene amplicons from ten diverse ocean environments demonstrated that, with the exception of the tropical oceanic regions, non-cyanobacterial sequences of proteobacterial origin dominated the diazotrophic

communities (Figure 4.1; Farnelid et al. 2011). These two wide-spread studies show that research into the full diversity of diazotrophs is urgently needed in order to understand their role in the input pathway of the marine nitrogen cycle. Following, the studies that identified non-cyanobacterial diazotrophs in the open ocean will be discussed.

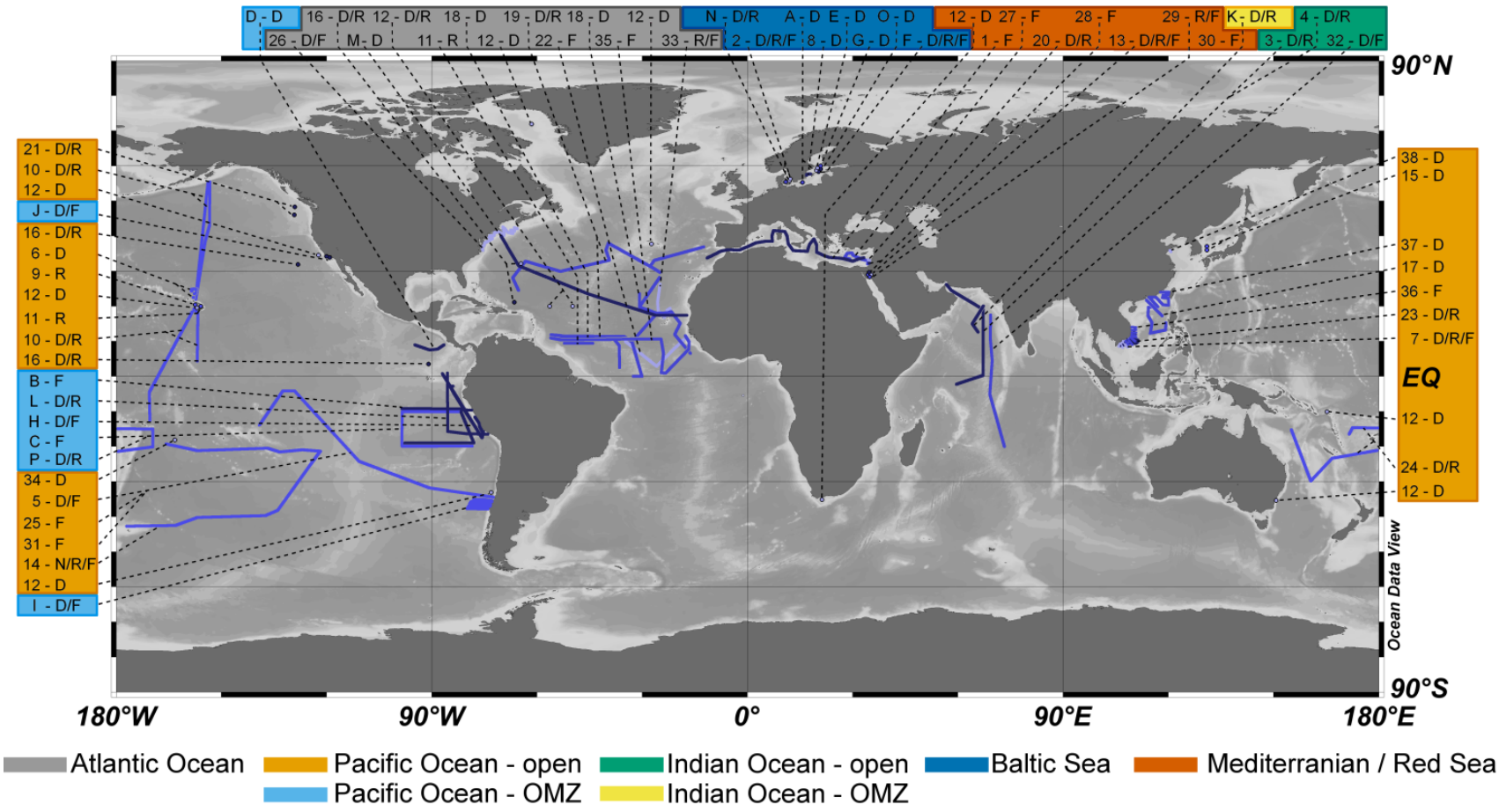

D - D
16 - D/R 12 - D/R 18 - D 19 - D/R 18 - D 12 - D
26 - D/F M - D 11 - R 12 - D 22 - F 35 - F 33 - R/F
N - D/R A - D E - D O - D
2 - D/R/F 8 - D G - D F - D/R/F
12 - D 27 - F 28 - F 29 - R/F
1 - F 20 - D/R 13 - D/R/F 30 - F
K - D/R 4 - D/R
3 - D/R 32 - D/F
90°N
90°S
180°W
90°W
0°
90°E
180°E
21 - D/R
10 - D/R
12 - D
J - D/F
16 - D/R
6 - D
9 - R
12 - D
11 - R
10 - D/R
16 - D/R
B - F
L - D/R
H - D/F
C - F
P - D/R
34 - D
5 - D/F
25 - F
31 - F
14 - N/R/F
12 - D
I - D/F
38 - D
15 - D
37 - D
17 - D
36 - F
23 - D/R
7 - D/R/F
EQ
12 - D
24 - D/R
12 - D
Ocean Data View
Atlantic Ocean
Pacific Ocean - open
Pacific Ocean - OMZ
Indian Ocean - open
Indian Ocean - OMZ
Baltic Sea
Mediterranian / Red Sea

Figure 4.1: Study sites that identified non-cyanobacterial diazotrophs or N_2 fixation rates associated with non-cyanobacterial diazotrophs.

The approximate research cruise tracks of studies summarized in Table 4.1 and Table 4.2 are indicated and labeled according to Table 4.1 and Table 4.2. These studies detected non-cyanobacterial diazotrophs or measured N_2 fixation in areas of non-cyanobacterial dominance. Studies labeled with numbers were performed in oxic waters only, whereas alphabetical labeled studies included samples from low O_2 water masses. Analysis of DNA, RNA or N_2 fixation is indicated with “D”, “R” or “F” respectively. Shades of blue represent maximum sampling depth: light blue – surface, medium blue – up to 200 m depth, dark blue – below 200 m depth. Due to crowding of the figure, extensive cruise transects by Langlois et al. (2015) were not included. The map was created using ocean data view (Schlitzer 2015).

4.3.1.1. Atlantic Ocean

Non-cyanobacterial diazotrophs have been detected in the Atlantic Ocean with the onset of *nifH* diversity studies (Table 4.1, Table 4.2 and Figure 4.2). In addition to extensive work on two clades of gamma-proteobacteria detected initially in clone libraries and thereafter using qPCR (Figure 4.1 and Figure 4.2; Langlois et al. 2015, 2008, 2005), Hewson et al. (2007) reported high phylogenetic *nifH* diversity and many heterotrophic *nifH* gene sequences in the eastern Atlantic Ocean. Turk et al. (2011) found that gamma-proteobacterial *nifH* expression was lower overall than that of the cyanobacteria in the eastern North Atlantic; however, in the western North Atlantic, extremely low cyanobacterial counts led to the conclusion that measured N_2 fixation rates must be attributed to heterotrophic diazotrophs (Mulholland et al. 2012).

Table 4.1: Studies reporting on non-cyanobacterial diazotrophs in the open ocean.

Label in Figure 4.1	Authors	Study Site	DNA/RNA/ fixation	Methods	Depth	Dominant Organisms	Conclusions
1	Benavides et al. 2016	Mediterranean	fixation	-	60 – 2000 m	-	Fixation in the aphotic zone correlated with organic material, but was negligible compared to overall nitrogen inputs
2	Bentzon-Tilia et al. 2015	Baltic Sea	DNA RNA fixation	Illumina TaqMan	9 + 35 m	*Pseudomonas stutzeri* in DNA and RNA, *Geobacter*/*Pelobacter*, cluster III, *Candidatus* A. thalassa, *Anabaena*	Monthly sampling 2012, Fixation peaked in spring and was higher in summer than in winter
3	Bird et al. 2005	Indian Ocean	DNA RNA	Sanger Sequencing, TaqMan	5 – 300 m	gamma-Proteobacterium, *Trichodesmium*	*Trichodesmium* only in the surface, deep dominated by gamma-proteobacteria
4	Bird et al. 2013	Indian Ocean	DNA RNA	Sanger Sequencing, after splitting by FACS	1 – 150 m	*Cyanothece*, *Crocosphaera*, gamma-Proteobacteria	Oligotrophic samples most diverse
5	Bombar et al. 2011	South China Sea	DNA RNA fixation	Sanger Sequencing TaqMan	surface	*Trichodesmium spp.*, *Richelia* sp., UCYN B, UCYN C, gamma-Proteobacteria, cluster III	Populations along the salinity and N:P ratio of the Mekong river plume. Oceanic+plume: *Trichodesmium* and *Richelia* (highest fixation), oceanic: UCYN B, UCYN C and gamma-proteobacteria
6	Bombar et al. 2013	North Pacific gyre, 2 cruises	DNA	Sanger Sequencing after splitting by FACS	5 – 175 m	*Candidatus* A. thalassa, *Trichodesmium, Lyngbya lagerheimii*, *Anabaena variabilis,* gamma-Proteobacterium	non-cyanobacterial abundances much lower than cyanobacterial, some non-cyanobacteria in the larger size fraction might be symbionts
7	Bonnet et al. 2008	South Pacific gyre	DNA fixation	Sanger Sequencing	30 m	*Vibrio diazotrophicus,* Proteobacteria, *Trichodesmium*, *Candidatus* A. thalassa	Cyanobacteria only detected at two stations in the gyre, N_2 fixation rates below the detection limit even after nutrient additions, outside the gyre higher cyanobacterial counts
8	Boström et al. 2007	Baltic Sea	-	Isolation	Deep suboxic water	closest cultured: *Pseudomonas stutzeri, Raoultella ornithinolytica*	Successful cultivation from low O_2 regime
9	Church et al. 2005	North Pacific	RNA	Sanger Sequencing, TaqMan	0 – 175 m	Cyanobacteria, gamma-Proteobacteria	Assessment of daily cycles of *nifH* expression: cyanobacteria have clear daily expression patterns of *nifH*, whereas the gamma-proteobacterium does not

10	Church et al. 2009	Pacific	DNA RNA	Sanger Sequencing, TaqMan	0 – 200 m	*Candidatus* A. thalassa, UCYN B, *Trichodesmium*, *Richelia* sp., Proteobacteria	Gamma-proteobacterial abundance low compared to cyanobacteria
11	Falcon et al. 2004	oligotrophic North Atlantic, North Pacific	RNA day and night sampling	Sanger Sequencing	euphotic zone	Cyanobacteria, alpha-Proteobacteria	North Pacific more unicellular cyanobacteria than North Atlantic, deeper samples more alpha-proteobacteria, more *nifH* transcripts at night
12	Farnelid et al. 2011a	10 Stations at multiple oceanic sites	DNA RNA	454 Sequencing	Surface	Proteobacteria, Cyanobacteria in tropical ocean	Proteobacteria dominated the temperate oceans, diversity in DNA and RNA did not correlate
13	Foster et al. 2009	Gulf of Aqaba	DNA RNA fixation	Sanger sequencing, TaqMan	2 – 5 m	*Trichodesmium*, Proteobacteria, *Candidatus* A. thalassa, *Crocosphaera*	*Trichodesmium* more abundant, Proteobacteria more evenly distributed
14	Halm et al. 2012	South Pacific Gyre	DNA RNA fixation	Sanger Sequencing, TaqMan	0 – 200 m	gamma-Proteobacteria, *Candidatus* A. thalassa	N_2 fixation rates were lowest in the ultra-oligotrophic centre of the gyre
15	Hashimoto et al. 2011	Japan Sea	DNA	Sanger Sequencing	0 – 150 m	Proteobacteria, *Candidatus* A. thalassa	No TaqMan to validate PCR amplified results
16	Hewson et al. 2007	Sargasso Sea, tropical Atlantic, Pacific	DNA RNA	Sanger Sequencing, TaqMan	Sargasso: surface, other stations: 0-6000 m	Surface: gamma-Proteobacteria, *Trichodesmium*, *Crocospheara*, Deep: Proteobacteria	North Pacific more diverse then North Atlantic, intermediate water depths more diverse than surface and deep
17	Kong et al. 2011	Western Pacific	DNA	Sanger Sequencing, SYBR qPCR	0 – 150 m	Proteobacteria, *Trichodesmium*, unicellular cyanobacteria, cluster III	No qPCR for heterotrophic phylotypes, no cyanobacteria in river plume and in winter (comparison of consecutive summer and winter)
18	Langlois et al. 2005	Tropical Atlantic	DNA	Sanger Sequencing	0 – 100 m	*Trichodesmium*, *Candidatus* A. thalassa, gamma-Proteobacteria, cluster III	Cyanobacteria dominated, gamma proteobacteria associated with colder and deeper water
19	Langlois et al. 2008	North Atlantic	DNA RNA	TaqMan	5 – 120 m	Cyanobacteria, gamma-Proteobacteria, cluster III	Cyanobacteria dominated the surface region
	Langlois et al. 2015	495 stations, 1000 samples, 15 cruises, multiple oceans	DNA RNA	TaqMan for Gamma A (AY896371)	Surface	Gamma A found in 67% of DNA samples, highest in tropical North Atlantic and subtropical North Pacific	Transcript counts were higher than DNA counts, higher counts at high temperatures, oxygenated surface waters, low nutrients and N deficit
20	Man-Aharonovich et al. 2007	Eastern Mediterranean Sea	DNA RNA	Sanger Sequencing	5 m	Proteobacteria, *Cyanothece* and *Candidatus* A. thalassa, methanogenic archaea, cluster III	extremely high N:P ratios, proteobacteria dominated, hardly any overlap of DNA and RNA, particle associated anaerobes

21	Mehta et al. 2003	Hydrothermal vents in North Pacific	DNA	Sanger Sequencing	Deep ocean	alpha-, gamma-Proteobacteria, Archaea, anaerobic bacteria (cluster III)	*nifH* diversity was highest in low DIN environment of vents, diversity was much lower in surrounding deep ocean water
22	Mills et al. 2004	Tropical North Atlantic	fixation	-	0 – 3 m	*Trichodesmium*	nutrient addition experiments: Diazotrophs were P and Fe limited
23	Moisander et al. 2008	South China Sea	DNA RNA	Sanger Sequencing	0 – 1700 m	*Trichodesmium*, alpha-, beta-, gamma- and delta-Proteobacteria	Cyanobacteria dominated in the surface
24	Moisander et al. 2014	eastern Australian coast	DNA, RNA at midday and midnight	Sanger sequencing, TaqMan for gamma-proteobacterium (g-24774A11)	0 - 175 m	*Candidatus* A. thalassa, *Trichodesmium spp.* and *Crocosphaera*	g-24774A11 was more evenly spread and had a higher total occurrence than the cyanobacterial phylotypes, potentially 26% of total fixation
25	Moutin et al. 2007	South Pacific gyre	fixation	-	0 – 200 m	-	Fixation lower in the gyre compared to stations outside the gyre
26	Mulholland et al. 2012	North Atlantic, 3 cruises in different seasons	DNA fixation	TaqMan	Surface	*Candidatus* A. thalassa, *Trichodesmium*, *Richelia*	Fixation higher in surface in summer, proteobacteria high in regions were cyanobacteria were low, but fixation still high
27	Rahav et al. 2013a	Eastern Mediterranean	fixation	-	0 – 150 m	Heterotrophic organisms	Nutrients, N_2 fixation and primary productivity increased from east to west
28	Rahav et al. 2013b	Gulf of Aqabe Levantine Basin	fixation	-	Aphotic zone	Heterotrophic organisms	Fixation higher in stratified summer period
29	Rahav et al. 2013c	Gulf of Aqabe	RNA fixation	Metatranscriptomics	60 – 130 m	Methanosarcinales, delta-Proteobacteria, Chlorobi	Fixation in stratified summer vs. mixed winter
30	Rahav et al. 2015	Gulf of Aqabe	fixation	-	10 – 160 m	-	N_2 fixation was highest during *Trichodesmium* bloom, rates increased after P addition and correlated with bacterial production, indicating heterotrophs contributed to fixing N_2
31	Raimbault et al. 2007	South Pacific gyre	fixation	-	0 – 200 m	-	N_2 fixation lower in the gyre compared to stations outside the gyre
32	Shiozaki et al. 2014	Arabian Sea, Equatorial Indian Ocean,	DNA fixation	Sanger sequencing	0 – 200 m	*Trichodesmium, Crocosphaera*, *Vibrio diazotrophicus*, alpha-, beta- and gamma-Proteobacteria	N_2 fixation highest at 12 – 25 m depth, and much higher in Arabian Sea than equatorial Indian Ocean, gamma-proteobacterium same counts as *Trichodesmium*
33	Turk et al. 2011	North Atlantic	RNA Fixation	Sanger Sequencing, TaqMan	Surface	*Candidatus* A. thalassa, *Trichodesmium*, Proteobacteria	PCR bias of gamma-proteobacteria, gamma-proteobacterial transcripts highest at one out of six stations

34	Turk-Kobo 2015	Noumea lagoon	DNA	mesocosm experiments: DIP addition to stimulate diazotrophs	surface	before DIP: Het-1, Het-2, UCYN A1 and A2 after DIP: Het-1 and UCYN C	Community progressed through different DIP concentrations
35	Voss 2004	Tropical North Atlantic	fixation	-	0-100 m	-	Rates were highest in surface waters close to the African coast
36	Voss et al. 2007	South China Sea	fixation	-	0 – 80 m	-	Highest rates during summer (upwelling on the coast) and at the river plume
37	Xiao 2015	South China Sea	DNA	454 pyro sequencing	0 – 200 m	gamma-Proteobacterium, *Trichodesmium*,	71% of reads belonged to 3 OTUs, community changed with depth
38	Zhang et al. 2015	South China Sea	DNA	Sanger Sequencing SYBR qPCR	0 – 30 m	Proteobacteria	Compare macroalgal canopies covered and not-covered areas; *Desulfovibrio* species in covered and *Vibrio* species in uncovered areas

Table 4.2: Studies reporting on non-cyanobacterial diazotrophs in O_2-deficient marine environments.

Label in Figure 4.1	Authors	Study Site	DNA/RNA/ fixation	Method	Depth	Dominant Organism	Conclusions
A	Bentzon-Tilia et al. 2014	Baltic Sea	DNA	Sager Sequencing, Cultivation	5 – 250 m	Sulfur and Sulfate reducer, cluster III	highest diversity at chemocline, successful isolation of *Pseudomonas stutzeri* species
B	Bonnet et al. 2013	OMZ off the Peruvian coast	fixation	-	0 – 2000 m	alpha- and gamma-Proteobacteria, sulfate reducing delta-Proteobacteria	Fixation rates highest in oxycline and OMZ core, but much lower compared to cyanobacteria
C	Dekaeze-macker et al. 2013	OMZ off the Peruvian coast	fixation	-	0 – 200 m	-	Fixation Rates lower in the HNLC region than in LNLC, P in excess
D	Cheung et al. 2016	Costa Rica Dome OMZ	DNA	Illumina	200 – 1000 m	*Vibrio diazotrophicus*, *Methylocella palustris*, Proteobacteria, cluster III	Proteobacteria dominated, phylotype distribution was governed by environment, diversity higher below the surface
E	Farnelid et al. 2009	Baltic Sea	DNA	Sanger Sequencing	3 m	Proteobacteria, Cyanobacteria, few Custer II and III	Time series April – October: diverse and changing community
F	Farnelid et al. 2013	Baltic Sea	DNA RNA fixation	454 pyro sequencing	5 – 233 m	*Nodularia*, alpha-, beta- and gamma-Proteobacteria, cluster III	RNA less diverse than DNA, highest diversity at chemocline, N_2 fixation at all depths
G	Farnelid et al. 2014	Baltic Sea	DNA	Sanger Sequencing, Cultivation	3 + 20 m	*Desulfovibrio*, *Burkholderia vietnamiensis*, *Nodularia* sp., gamma-Proteobacteria	Successful cultivation
H	Fernandez et al. 2011	OMZ off the Peruvian coast	DNA fixation	Sanger Sequencing	0 – 400 m	alpha-, beta-, gamma-, delta -Proteobacteria, Methanobacteria, Firmicutes	N_2 fixation detected, but no cyanobacteria present
I	Fernandez et al. 2015	OMZ off the Peruvian coast (3 cruises and COPAS)	DNA fixation	Sanger Sequencing	0 – 50 m	alpha-, beta- and gamma-Proteobacteria	Fixation rates significantly higher than previously measured in that region, Low N:P

J	Hamersley et al. 2011	Four year sampling at SPOTS and SMBO	DNA fixation	Sanger Sequencing	0 – 900 m	*Candidatus* A. thalassa, *Richelia* sp., alpha- and gamma –Proteobacterium, sulfate reducers, cluster III	Low N:P ratios, N_2 fixation correlated with SST and highest in the surface
K	Jayakumar et al. 2012	Arabian Sea upwelling	DNA RNA	Sanger Sequencing	Oxic – OMZ core	alpha-, gamma-Proteobacteria, sulfate reducers, cluster III	No cyanobacteria
L	Loescher et al. 2014	OMZ off the Peruvian coast	DNA RNA	Sanger Sequencing	Oxic – core of OMZ	Proteobacteria, sulfate-reducing bacteria, *Crocosphaera*	Phylotypes inhabited different niches: Shelf, high nutrient, open ocean; fixation rates highest during sulfidic event
M	Ratten et al. 2015	North Atlantic upwelling	DNA	TaqMan	0 – 200 m	Gamma A	Gamma A dominated in low O_2 upwelling region
N	Severin et al. 2015	Baltic Sea	DNA RNA	Illumina, TaqMan	Surface	Proteobacteria	Lowering O_2 and increasing glucose availability, changed the diazotrophic community: overall, the ratio of potentially fixing : actually fixing increased
O	Thureborn et al. 2013	Baltic Sea	DNA	Metagenomics	Oxic to anoxic water	sulphate-reducing delta-Proteobacteria	Nif genes detected in suboxic and sediment samples
P	Turk-Kubo et al. 2014	OMZ off the Peruvian coast	DNA RNA	Sanger sequencing	0 – 200 m	gamma-proteobacteria, cluster I – III	Higher diversity in LNLC samples, gamma-proteobacteria cannot account for fixation rates measured

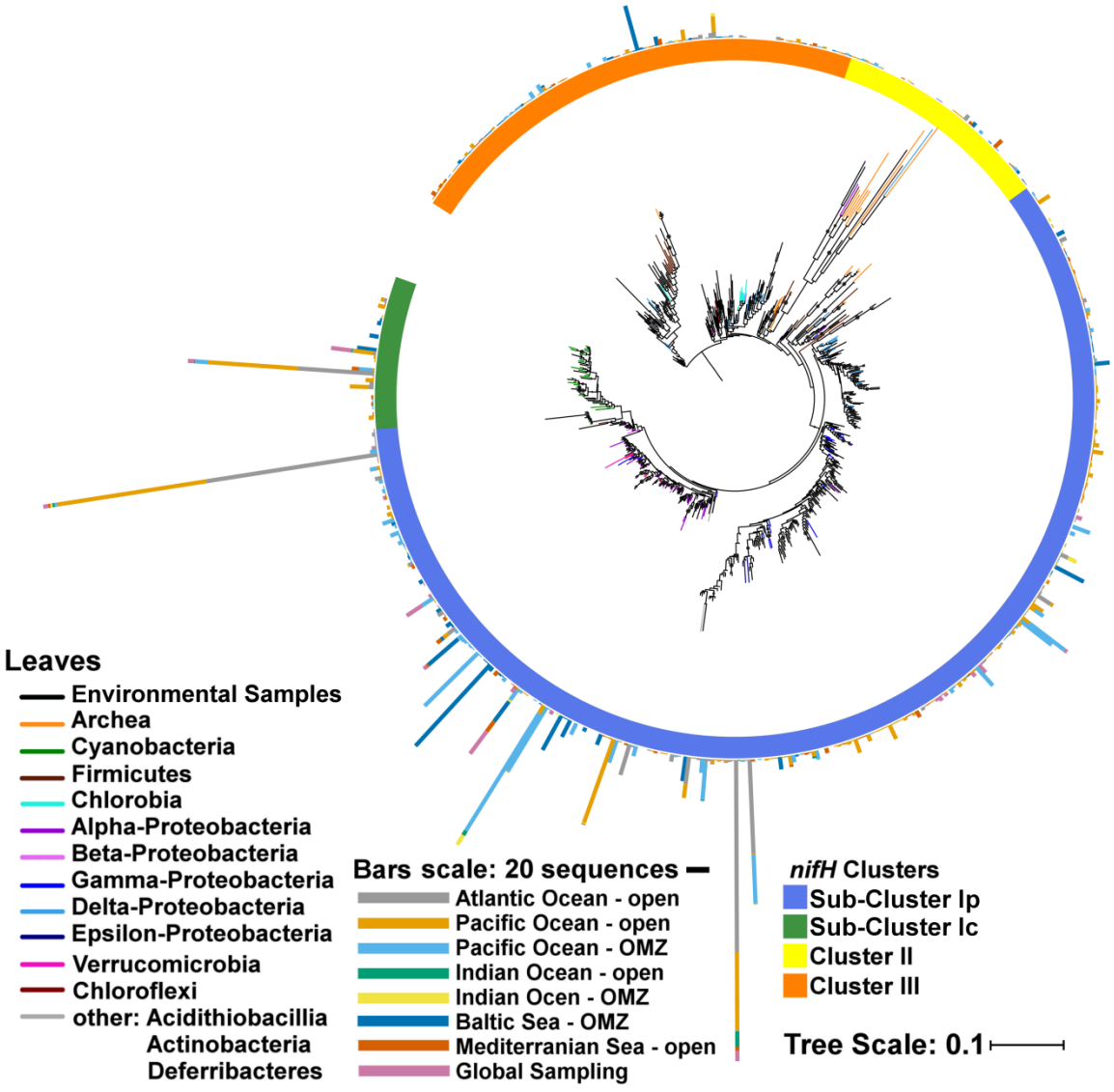


Figure 4.2: Phylogenetic analysis of extracted and clustered *nifH* sequences.

The *nifH* sequences from studies that identified non-cyanobacterial diazotrophs were extracted from NCBI along with their most closely related reference genomes. The sequences were clustered at 96% identity using CD-HIT (Li and

Godzik 2006). The phylogenetic affiliation was inferred by maximum likelihood tree building based on the GTR-GAMMA model of codon-aligned *nifH* sequences in RAxML (MAFFT v. 7; Yamada et al. 2016; Stamakatis 2014; Katoh et al. 2002). Bootstrap values were calculated from 100 replicates and values >50% are shown as dots. The tree was displayed with branch lengths showing the number of substitutions per site Leaves of reference genomes are coloured according to their taxonomy. Black leaves indicate environmental sequences. The outer bar graph indicate the number of sequences present in each OTU cluster from all studies combined. Bars are coloured based on the location of their detection and clusters are assigned according to Zehr et al. (2003).

4.3.1.2. Pacific Ocean

In the western South Pacific, surface waters off the eastern Australian coast and in New Caledonia were dominated by cyanobacterial phylotypes *Richelia* sp. and *Candidatus* Atelocyanobacterium thalassa (UCYN A; Figure 4.2; Turk-Kubo et al. 2015; Moisander et al. 2014). Also near New Caledonia, mesocosm experiments were dominated by cyanobacterial diazotrophs, but dissolved organic phosphate (DIP) amendments led to an increase in the relative abundance of gamma-proteobacterial *nifH* counts, possibly caused by increased substrate for heterotrophs from enhanced primary productivity (Turk-Kubo et al. 2015). Independent observations of nocturnally-increased abundances of gamma-proteobacterial *nifH* transcripts also support a dependency on primary production (Moisander et al. 2014).

The diazotrophic communities in the North Pacific, and at HOT (Hawaii Ocean Time-Series) in particular, were dominated by the symbiotic *Candidatus* A. thalassa. However, proteobacterial and cluster III sequences have consistently been recovered from DNA and RNA in that region (Figure 4.2; Bombar et al.

2013; Church et al.2009; Zehr et al. 2008; Church et al. 2005). A high diversity of non-cyanobacterial *nifH* sequences from cluster I and III was observed in the South Pacific gyre, although a few *Candidatus* A. thalassa sequences were also detected (Halm et al 2012; Bonnet et al 2008). Active N_2 fixation by non-cyanobacterial organisms in the South Pacific gyre was supported by measurable, albeit low, N_2 fixation rates in the absence of high cyanobacterial *nifH* counts (Halm et al. 2012; Bonnet et al. 2008; Moutin et al. 2007; Raimbault et al.2007).

As expected, diazotrophic communities in deep hydrothermal vent sites in the North Pacific were composed exclusively of non-cyanobacterial *nifH* phylotypes from clusters I – III including alpha- and gamma-proteobacteria, anaerobic organisms from cluster II and a majority of diversity from cluster III (Mehta et al. 2003). The *nifH* diversity was highest at low DIN concentrations.

Overall, the Western North Pacific surface waters are dominated by cyanobacterial diazotrophs. Below the photic zone, diazotrophic diversity was higher and included mainly non-cyanobacterial organisms of clusters I and III (Figure 4.2). However, N_2 fixation rates for deep samples was not measured in any of the relevant studies.

4.3.1.3. Mediterranean and Red Sea

The eastern Mediterranean Sea is one of the most oligotrophic regions in the world; its deep waters display elevated dissolved N:P ratios compared to the

Redfield Ratio (25-28:1; Krom et al. 1991), which may stem from high N_2 fixation in the recent geological past (Pantoja et al. 2002; Bethoux et al 1992). The identification of non-cyanobacterial *nifH* in DNA and cDNA sequences from the eastern Mediterranean Sea has confirmed the presence of diverse *nifH* phylotypes belonging to clusters I, II and III (Man-Aharonovich et al. 2007) However, measurements of N_2 fixation in the Mediterranean Sea were low both in and below the photic zone in areas dominated by non-cyanobacterial diazotrophs (Benavides et al. 2016; Rahav et al. 2013a; Bonnet et al. 2011; Yogev et al. 2011).

Similar to the eastern Mediterranean sea, the typically oligotrophic stratified surface waters of the Gulf of Aqaba is replenished with nutrients during winter mixing (Manasrah et al. 2007; Fuller et al. 2005; Labiosa and Arrigo 2003). In the photic layer of the Gulf of Aqaba, N_2 fixation rates were low (Foster et al. 2009) and the correlation of N_2 fixation rates with bacterial production indicated that non-cyanobacterial diazotrophs were contributing to N_2 fixation in the photic layer (Rahav et al. 2015). In contrast, the Gulf of Aqaba showed significant N_2 fixation rates below the photic layer where non-cyanobacterial organisms dominated (Rahav et al. 2013b and c).

4.3.2. Coastal Seas

In the tropical South China Sea region of the western Pacific, actively-fixing cyanobacterial diazotrophs, especially *Trichodesmium*, dominated the diazotrophic community in surface waters. Diversity of non-cyanobacterial *nifH*

phylotypes increased with depth (Figure 4.2; Xiao et al. 2015; Kong et al. 2011; Moisander et al. 2008; Voss et al. 2006). Cyanobacterial *nifH* phylotypes were abundant during the summer, but not detected in winter months (Kong et al. 2011; Xiao et al. 2015; Bombar et al. 2011; Moisander at al. 2008). Within the same region, in the surface waters of the Mekong river plume, *Trichodesmium spp.* and *Richelia* sp. were dominant with only a few gamma-proteobacterial and cluster III sequences detected (Bombar et al. 2011). In the temperate Sea of Japan, proteobacteria dominated clone libraries; no *Trichodesmium spp.* was detected and *Candidatus* A. thalassa was not abundant (Hashimoto et al. 2011).

4.3.3. Oxygen deficient waters

On the eastern sides of the Atlantic and Pacific Oceans, and in the Arabian Sea, upwelling of deep nutrient-rich water to the sunlit surface supports rapid phytoplankton growth; this is followed by a severe drawdown of dissolved O_2 in subsurface waters as primary production sinks and decays, resulting in oxygen minimum zones (OMZ). OMZs are also observed in the Baltic Sea and the California Bight, where geographical barriers and a permanent halocline prevent water column mixing. Consequently, respiration at depth results in permanently O_2-deficient deep water, and low N:P ratios due to anaerobic respiration of nitrate (Berelson 1991; Conley et al. 2002). Although O_2-deficient regions make up less than 0.1% of total oceanic volume, it is estimated that anaerobic nitrogen loss pathways such as denitrification and anammox lead to the removal of 25 – 50% of oceanic DIN resulting in a high P* value, which indicates excess P relative to

fixed N throughout OMZs (Gruber 2008; Codispoti et al. 2007, 2001; Hamersley et al. 2007; Ingall and Jahnke 1994). Excess P, together with high Fe concentration measured in OMZs, suggest that O_2-deficient regions provide an undiscovered niche for diazotrophs, linking N loss and gain processes spatially (Deutsch et al. 2007). The predicted expansion of OMZs with ocean warming calls for an assessment of these areas as potential hotspots for N_2 fixation (Stramma et al. 2008, 2012).

4.3.3.1. Atlantic Ocean

There is conflicting evidence relating to N_2 fixation in the Benguela upwelling system of the South Atlantic. Sohm et al. (2011a) and Subramaniam et al. (2013) detected low N_2 fixation rates that could not be attributed to cyanobacteria or Gamma A (Sohm et al. 2011a). Others failed to detect N_2 fixation during seven cruises through the Benguela upwelling and were not successful in attempts to induce N_2 fixation through nutrient addition experiments (Wasmund et al. 2015). This led to the conclusion that there was a general absence of diazotrophs in this system and suggest that previously measured N_2 fixation rates were from diazotrophs introduced into the region through lateral transport of warmer water masses. Gamma A was found in the OMZ off the northwestern African coast, although N_2 fixation rates were not measured in that study (Ratten et al. 2015, Chapter 2).

4.3.3.2. Pacific Ocean

The most pronounced OMZs off the Peruvian and Chilean coasts are permanent oceanographic features that exhibit an annual cycle of suppressed or absent upwelling in autumn and winter and active upwelling in spring and summer: the active upwelling season leads to anoxia and occasional sulfidic events (Schunck et al. 2013; Thamdrup et al. 2012; Sobarzo et al. 2007; Daneril et al. 2000). Diazotrophs in this OMZ have been sampled frequently since Deutsch et al. (2007) proposed a potential diazotrophic niche in these denitrified water masses (Figure 4.1). Measurements of N_2 fixation rates spanning the water column from the oxygenated surface over the oxycline into the core of the OMZs showed active N_2 fixation at all depths (up to 24.8 nmol N L^{-1} d^{-1} during a sulfidic event when O_2 is completely depleted; Loescher et al. 2014; Bonnet et al. 2013; Dekaezemacker et al. 2013; Fernandez et al 2011). Although average N_2 fixation rates below the photic zone in the OMZ were generally low, the large volume of the aphotic zone relative to the photic zone implies that aphotic N_2 fixation rates could make up over 90% of the total N_2 fixation in the OMZ water column, thereby compensating for at least 11% of nitrogen lost through anaerobic respiration (Bonnet et al. 2013; Dekaezemacker et al. 2013).

The *nifH* phylotypes recovered from the South American OMZ were closely related to *nifH* phylotypes previously identified in the Pacific and Indian Ocean and were predominantly of the non-cyanobacterial types from clusters I and III (Fernandez et al. 2015; Loescher et al. 2014; Turk-Kubo et al. 2014; Halm et al. 2012; Farnelid et al. 2011a ; Fernandez et al. 2011). Environmental factors

shaping the spatial distribution of these non-cyanobacterial *nifH* phylotypes included dissolved O_2, nitrite, and temperature (Loescher et al. 2014; Cheung et al. 2016). Further south, an 18 month sampling regime at COPAS (Center for Oceanographic Research in the Eastern South Pacific) demonstrated a recurring seasonal pattern of N_2 fixation; peak N_2 fixation rates occurred at the end of summer under highly-productive, low N:P ratios, and suboxic conditions (Fernandez et al. 2015). The authors thus proposed a C:N:P regulation of N_2 fixation.

In all phylogenetic studies of the South Pacific OMZ, cyanobacterial diazotrophs were found only in low abundance throughout the surface ocean and not at all at depth (Figure 4.2). The presence of moderate N_2 fixation rates in the absence of cyanobacteria both in the surface and below the photic layer supports the hypothesis that some non-cyanobacterial diazotrophs were performing N_2 fixation (Fernandez et al. 2015; Loescher et al. 2014; Bonnet et al. 2013; Fernandez et al. 2011).

4.3.3.3. Indian Ocean

The largest suboxic region in the world's ocean occurs in the Arabian Sea (Naqvi 2008). This area also experiences large amounts of dust deposition at the surface, which has been proposed to enhance diazotrophic activity by providing iron for the nitrogenase enzyme (Jickells et al. 2005). There are few studies on non-cyanobacterial diazotrophs in the Indian Ocean. Coinciding with highest N_2 fixation rates, the euphotic zone was found to be dominated by *Trichodesmium*,

Cyanothece and *Crocosphaera*, whereas the aphotic zone was more diverse (Shiozaki et al. 2014; Bird and Wyman 2013; Jayakumar et al. 2012; Bird et al. 2005). Active expression of gamma-proteobacterial *nifH* sequences (among them Gamma A) was demonstrated in the aphotic zone (Shiozaki et al. 2014; Bird and Wyman 2013; Jayakumar et al. 2012).

4.3.3.4. Anoxic Basins – California Bight and Baltic Sea

There are two oceanographic sampling stations in the California Bight: SPOTS (San Pedro Ocean Time Series) and SMBO (Santa Monica Bay Observatory). In a four-year time-series study at both stations, sea surface temperature correlated with N_2 fixation rates and *nifH* diversity (Hamersley et al. 2011). Fixation rates were considerably higher at the surface compared to intermediate and deep waters, possibly due to the presence of *Candidatus* A. thalassa and *Richelia* in the surface. At depth, only non-cyanobacterial sequences from cluster I and III were found (Figure 4.2).

The anoxic basins of the Baltic Sea have a salinity of 3 – 8 at the surface and 4 – 14 at the bottom (Janssen et al. 1999). Like oceanic regions, cyanobacterial blooms were found in the surface waters while an increasingly diverse non-cyanobacterial diazotrophic community was found at depth (Farnelid et al. 2013). Highest *nifH* diversity and abundances at the chemocline suggest that this area is the most suitable non-cyanobacterial niche; here, O_2 concentrations are sufficient for aerobic respiration and oxidative phosphorylation, but low enough to efficiently protect the nitrogenase from oxidation (Farnelid et al. 2013). Comparing *nifH*

transcript diversity to *nifH* DNA diversity showed that the potential diazotrophic community was much larger than the transcriptionally-active community (Farnelid et al. 2013). The surface waters also displayed a diverse and changing community in three surface time-series (Figure 4.2; Bentzon-Tilia et al. 2015; Farnelid et al. 2009). In two of these time-series, the dominating organism in both DNA and RNA samples was *Pseudomonas stutzeri* strain, recently isolated from the Baltic Sea (Bentzon-Tilia et al. 2015). Despite the absence of cyanobacterial *nifH* sequences, N_2 fixation rates were high from spring until autumn, at times reaching rates as high as those measured in open ocean surface waters (Bentzon-Tilia et al. 2015).

Attempts to cultivate non-cyanobacterial diazotrophs have been successful in the Baltic Sea (Bentzon-Tilia et al. 2015; Bentzon-Tilia et al. 2014; Farnelid et al. 2014). The isolated organisms belong to the species *Pseudomonas stutzeri* (gamma-proteobacterium), *Raoultella ornithinolytica* (gamma-proteobacterium) and *Rhodopseudomonas palustris* (alpha-proteobacterium; Bentzon-Tilia et al. 2015; Bentzon-Tilia et al. 2014; Farnelid et al. 2014). Their N_2 fixation and growth rates in pure culture were tightly regulated by the availability of glucose and O_2. In the presence of excess glucose and optimal O_2 conditions (165, 38 and 14 μmol O_2 L^{-1} respectively), N_2 fixation rates for each of the three organisms were high enough to contribute significantly to total marine N_2 fixation rates observed in these areas (Bentzon-Tilia et al. 2015). Differing optimal conditions for growth and N_2 fixation rates were also shown in an O_2 reduction and carbon addition experiment in bottle experiments from a nitrogen depleted Danish Fjord (Severin

et al. 2015). The fact that marine non-cyanobacterial diazotrophs are actively fixing N_2 and potentially reach rates that are significant compared to the surrounding community, cannot be disputed in the context of these studies.

There has been a widespread skepticism surrounding the contribution, if any, of non-cyanobacterial diazotrophs to marine N_2 fixation. This controversy stems primarily from reported potential contamination by exogenous *nifH* sequences from molecular biology reagents or soil-derived dust particles; putatively, these contaminating sequences would amplify more readily when diazotrophs are low in abundance or absent in natural microbial communities (Izquierdo et al. 2006; Goto et al. 2005; Zehr et al. 2003b). However, the proposal that non-cyanobacterial diazotrophs are widely distributed in the marine environment is founded on strong evidence, including the repeated detection of non-cyanobacterial diazotrophs despite varied sample collection and extraction methods, quantitation of both DNA sequences and actively transcribed *nifH* via qPCR assays, the measurements of N_2 fixation rates in the absence of cyanobacterial diazotrophs in waters with high DIN concentrations, and the isolation of non-cyanobacterial diazotrophs from marine environments; combined, these point to the significant presence and importance of these microorganisms in large oceanic regions. At present, their role in the marine nitrogen cycle, and in the environment in general, remains unclear and deserving of further exploration.

The existence of non-cyanobacterial diazotrophs in diverse oceanic habitats introduces new questions regarding the function and significance of diazotrophs, including two metabolic conundrums: Firstly, how can the nitrogenase be kept in

an anaerobic environment in the oxygenated ocean? And secondly, why would diazotrophs perform N_2 fixation, a highly energy-demanding process, in waters where fixed nitrogen is freely available in high concentrations?

4.4. Metabolism

N_2 fixation is estimated to have evolved before the oxygenation of the atmosphere by cyanobacteria, because the reactive site of the nitrogenase is highly susceptible to oxidation (Summons et al. 1999; Falkowski 1997). Consequently, diazotrophs have developed adaptive mechanisms to exclude O_2 from the proximity of the nitrogenase enzyme during periods of N_2 fixation. O_2 evasion mechanisms have been studied extensively in marine diazotrophic cyanobacteria, because the O_2 produced during photosynthesis could directly oxidize the nitrogenase at already ambient atmospheric O_2 concentrations (Zehr 2011; Berman-Frank et al. 2003). For example, in the chain forming cyanobacteria *Richelia*, the site of N_2 fixation is within differentiated cells (heterocysts) on either end of the chain. Heterocysts possess a thick cell wall that excludes O_2 and hence, provide a suitable environment for the nitrogenase (Haselkorn 2007; Jahson et al. 1995). This mechanism has also been found in terrestrial diazotrophs (Witty and Minchin 1998). *Candidatus* A. thalassa has developed a completely different mechanism: this species has lost the genes for proteins of photosystem II, which is the O_2-evolving complex of photosynthesis, along with other genes coding for proteins required during carbon fixation (Zehr et al. 2008). As *Candidatus* A. thalassa is a symbiont, its eukaryotic host,

Braarudosphaera bigelowii, supplies a carbon source and in return, *Candidatus* A. thalassa synthesises ammonium (Krupke et al. 2016; Thompson et al 2014; Krupke et al. 2013; Zehr et al. 2008). On the other hand, the unicellular cyanobacteria *Cyanothece* and *Crocosphaera* alternate between photosynthesis during the day and N_2 fixation at night (Bandyopadhyay et al. 2011; Mohr et al. 2010; Reddy et al. 1993), thereby avoiding the high O_2 concentrations evolved during photosynthesis. However, the organisms still have to deal with ambient O_2, which is especially high in the surface ocean.

Without established cultured representatives, it is difficult to assess the adaptive mechanisms for O_2 reduction developed by non-cyanobacterial diazotrophs inhabiting oxygenated waters. While non-cyanobacterial diazotrophs do not evolve oxygen, aerobic diazotrophs must balance the need to prevent oxidative damage of the nitrogenase complex with the respiratory oxygen required to produce ATP via oxidative phosphorylation. Anaerobic respiration can only provide sufficient energy for N_2 fixation when coupled with an alternate electron accesptor than oxygen (Madigan 1995).

O_2 evasion has been well studied in the free-living soil bacterium *Azotobacter vinelandii* (Giuffrè et al. 2014; Dixon and Kahn 2004). *Azotobacter* sp. have developed a vast array of mechanisms to deal with fluctuating O_2 concentrations. Such mechanisms include: conformational changes and protein complex protection (Schlesier et al. 2015; Moshiri et al. 1995), respiratory protection (Inomura et al. 2017; Paulus et al. 2012; Poole and Hill 1997; Juenemann et al. 1995; Kolonay et al. 1994), and synthesis of reducing equivalents (Thorneley and

Ashby 1989). Bacterial interactions such as cell-to-cell clumping, flocculation and the synthesis of O_2 reducing enzymes have also been shown to be involved in O_2 reduction in various bacterial species (Bentzon-Tilia et al. 2015; Bible et al. 2015; Dingler et al. 1988; Dingler and Oelze 1987). It has also been proposed that intense respiration on organic aggregates can lead to significantly lower O_2 concentrations and a drawdown of available DIN through enzymatic hydrolysis uncoupled from uptake, which could favour diazotrophic metabolism (Ploug and Buchholz 1997; Smith et al. 1992; Paerl, Prufert 1987). This hypothesis is supported by the detection of strictly anaerobic cluster III *nifH* sequences throughout the open ocean (Rieman et al. 2010; Moisander et al. 2008; Zehr et al. 2003a; van der Maarel et al. 1999; Marty 1993; Sieburth 1987). However, these adaptations cannot be applied to all non-cyanobacterial diazotrophs living in the soil or the marine environment. Thus, whether terrestrial O_2 evasion mechanisms apply to marine non-cyanobacterial diazotrophs is unclear.

Although diazotrophs found in the oceanic OMZs have potentially solved the problem of having to deal with high ambient O_2 concentrations, they provide another conundrum: They, as well as diazotrophs found in low- O_2 hydrothermal vent communities and in estuarine, euphotic, mesopelagic, and benthic environments, show the potential to fix N_2 despite high ambient DIN concentrations (Knapp 2015; Voss et al. 2006). Initially, N_2 fixation was thought to be suppressed at DIN concentrations of greater than 1 µM, but it was shown that the uptake of nitrate is not much more energy-favourable than N_2 fixation in low O_2 concentrations around the nitrogenase; maintainance of low O_2 concentrations

is what makes N_2 fixation so energy-demanding (Großkopf and LaRoche 2012; Falkowski 1983). Both culture and field studies have shown a more complex regulation of N_2 fixation in regards to DIN concentrations present. In the field, N_2 fixation was observed at a range of 5 – 20 µM NO_3^- (Fernandez et al. 2011; Sohm et al. 2011b; Voss et al. 2004). Proliferation and regulation of N_2 fixation has been proposed to depend not only on DIN present, but also on iron and organic carbon accessibility. A twenty-fold higher iron-requirement has previously been shown to influence the distribution of diazotrophs with the nitrogenase requiring at least 20 iron atoms (Berman-Frank et al. 2007; Kustka et al. 2003; Berman-Frank et al. 2001). Furthermore, N_2 fixation might be regulated by nutrient ratios including low N:P or N:P:C ratios (Severin et al. 2016; Fernandez et al. 2015; Turk-Kubo et al. 2015; Moisander et al. 2014; Bombar et al. 2011), which would support the hypothesis that OMZs provide niches for diazotrophs (Deutsch et al. 2007).

4.5. Phylogeny and metabolic potential of non-cyanobacterial diazotrophs

To investigate the phylogenetic diversity of non-cyanobacterial diazotrophs, published *nifH* sequences were extracted from NCBI (ca. 3700 sequences; not including results from high-throughput sequencing technology with < 300 bp read length) along with their closely related reference genomes (132 genomes, of which 42 were isolated or found in aquatic environments; Supplemental Table 4 and Supplemental Table 5). The *nifH* sequences were clustered at 96% similarity

(OTUs) and a maximum likelihood tree was inferred using the GTR-GAMMA model in RAxML (Figure 4.2; Stamatakis et al. 2014). The uncultured *nifH* sequences clustered with reference genomes from the taxonomic groups of Archaea, Acidithiobacillia, Actinobacteria, Chlorobi, Chloroflexi, Cyanobacteria, Deferribacteres, Firmicutes, Proteobacteria and Verrucomicrobia, dividing the sequences into four clusters (I – IV) with cluster I being subdivided into sub-cluster Ic (dominated by cyanobacterial diazotrophs) and sub-cluster Ip (dominated by proteobacterial sequences; Zehr et al. 2003a). Sub-cluster Ip displayed the highest number of isolated sequences along with the greatest diversity, as was also seen in Farnelid et al.'s (2011a) global ocean sampling. Most *nifH* sequences of non-cyanobacterial diazotrophs were identified in the Pacific Ocean; however, no specific distribution patterns between the oceanic basins was observed (Figure 4.2). Cluster IV resembles an enzyme paralogue to the nbitrogenase, that does not contribute to nitrogen fixation, but is occasionally amplified during *nifH* PCR.

A phylogenetic tree was also built from 16S rRNA gene sequences of the reference genomes for phylogenetic comparison to their *nifH* sequences (Figure 4.3). Comparing the *nifH* gene tree to the 16S rRNA gene tree showed that *nifH* phylogeny is not always equivalent to 16S rRNA gene phylogeny (Figure 4.3). This was mainly manifested in Proteobacteria, Archaea and Firmicutes, suggesting that lateral gene transfer (LGT) of the *nifH* gene has occurred among these groups. In contrast, Cyanobacteria and Chlorobi *nifH* sequences each form separate phylogenetic clades that may point to vertical inheritance within these

groups. Of note, relationships among the major *nifH* gene clusters cannot be completely resolved and bootstrap support values are low, even when using full length *nifH* sequences or phylogenetic analysis of protein alignments (<50%, data not shown). When assigning taxonomy to unknown *nifH* sequences based on phylogeny, the possibility of lateral gene transfer and phylogenetic uncertainty must be taken into consideration to avoid false conclusions. However, because the following metabolic diversity analysis was carried out on completely sequenced and classified organisms, taxonomic assignment is not a problem in this case.

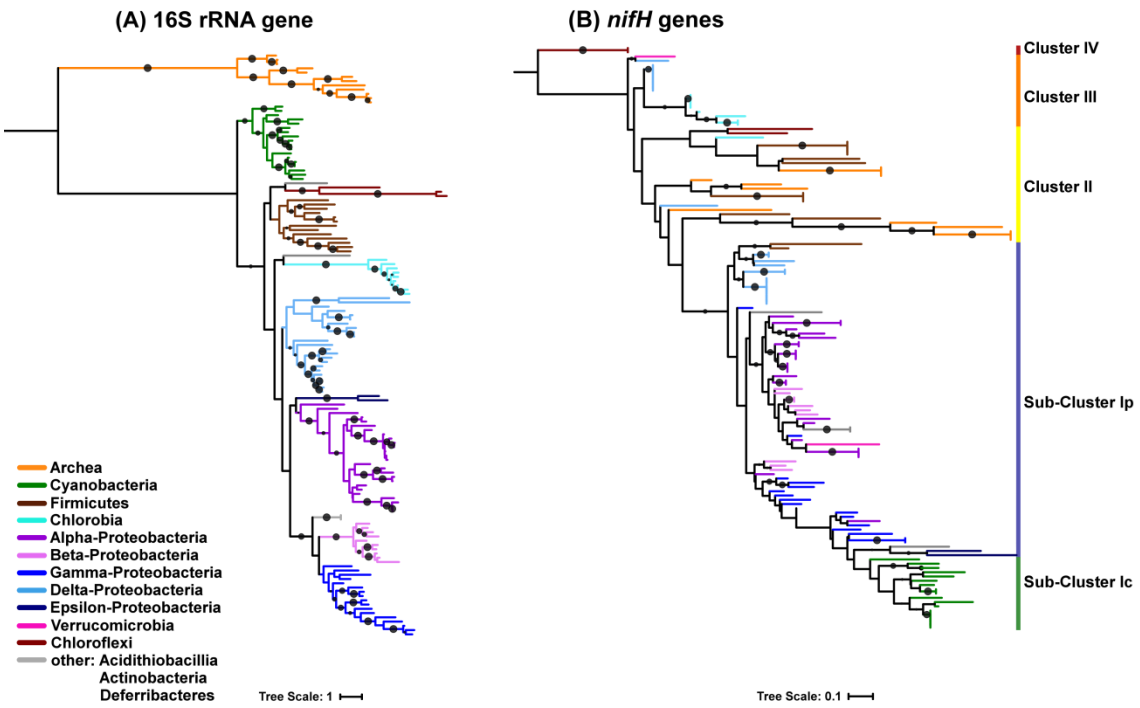


Figure 4.3: Phylogenetic analysis of the full-length 16S ribosomal RNA and *nifH* genes from reference genomes.

The phylogenetic associations of the full-length 16S ribosomal RNA genes (A) and the *nifH* genes (B) from reference genomes were inferred using the maximum likelihood method based on the GTR-GAMMA model after aligning nucleotide sequences of 16S rRNA and protein sequences of *nifH* genes (MAFFT v. 7; Yamada et al. 2016; Stamakatis 2014; Katoh et al. 2002). Bootstrap values were calculated from 100 tree replicates and values >50% are shown as dots. The tree was displayed with branch lengths showing the number of substitutions per site and clusters are assigned according to Zehr et al. (2003).

To garner essential information on the metabolic potential in non-cyanobacterial diazotrophs, the extracted reference genomes were annotated using FROMP and SEED (Desai et al. 2013; Overbeek et al. 2005). The presence or absence of 167 metabolic pathways involved with nutrient cycling as annotated by SEED was noted (carbon, aromatic compounds, nitrogen, phosphate, sulfur and iron;

Overbeek et al. 2005). The clustering algorithm of R's package *gplots* was applied to metabolic pathways that were found in more than three reference genomes. There was an association between taxonomy and metabolic pathways and hence, significant associations of metabolic pathways with taxonomic groups were determined statistically in R using the multilevel pattern analysis *indicspecies* (Figure 4.4, Table 4.3 and Table 4.4; R core team 2015; Warnes et al. 2013; Caceres and Legendre 2009).

The KEGG pathways that were identified via FROMP were used to gauge the overall differences in metabolic pathways among the reference genomes. An ANOSIM (Analysis of Similarities) test was performed in PRIMER. Overall, the metabolism of each taxon was significantly different from others (R statistic = 0.485, significance level = 0.1%; Clarke and Gorley 2006). Significant pairwise differences between taxonomic groups are recorded in Supplemental Table 6. The metabolism of the taxonomic group of Archaea was the most segregated from the other groups (Supplemental Table 6). Based on only 16S rRNA genes, this segregation was also seen in phylogenetic relationships (Figure 4.3).

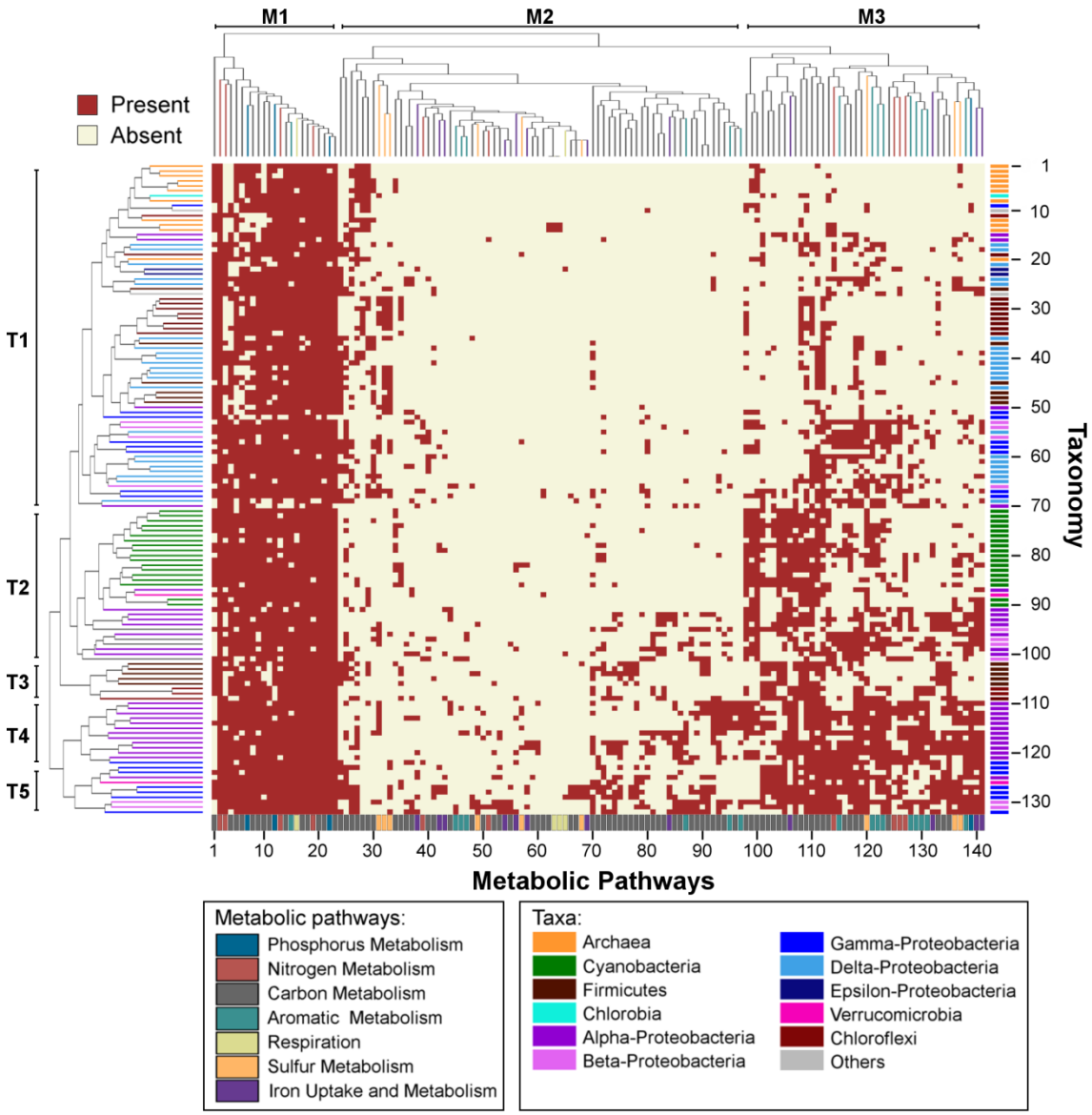


Figure 4.4: Metabolic potential present in marine-related diazotrophs.

The presence or absence of 167 pathways for the metabolism of carbon, aromatic compounds, nitrogen, phosphate, sulfur and iron was determined through annotation of reference genomes in the SEED database (Overbeek et al. 2005). Clustering analysis was performed with 142 metabolic pathways that were present in at least 3 reference genomes using the R package *gplots* (Warnes et al. 2013; R core team 2015). Metabolic pathways and taxonomy are colour coded according to the legend. Organisms and metabolic pathways are listed in Supplemental Table 4 and Supplemental Table 5.

The clustering analysis showed that there are common metabolic pathways throughout the reference genomes (M1: pathways 1 – 23; Figure 4.4). They include the essential pathways of energy metabolism from carbon sources (Glycolysis and TCA cycle), mono- and polysaccharide metabolic pathways, and anaerobic respiration. Phosphorus uptake and metabolism, N_2 fixation, ammonium uptake and ammonification are also mostly ubiquitous pathways. The remaining pathways clustered into two groups, one of which contains metabolic pathways that are rare (M2: pathways 24 – 97; Figure 4.4) relative to the other cluster (M3: pathways 98 – 141; Figure 4.4). The occurrence of metabolic pathways in these two variable clusters aligned somewhat with taxonomic category. Reference genomes clustered into five groups that were dominated by certain taxa (Figure 4.4). Archaea, Chlorobi, delta- and epsilon-Proteobacteria shared more metabolic pathways (T1: taxa 1 – 70; Figure 4.4) than the group of Cyanobacteria and Alpha- and beta-Proteobacteria (T2: taxa 71 – 101), which was different from Firmicutes (T3: taxa 102 – 109), a second collection of alpha-Proteobacteria (T4: taxa 110 – 123) and finally gamma-Proteobacteria (T5: taxa 124 – 132). The first cluster, T1, was associated with the least number of metabolic pathways (see also Table 4.3 and Table 4.4). However, 1C metabolism (formaldehyde assimilation and methanogenesis), and sulfur oxidation pathways, were more common in this cluster than in others. The cluster of Cyanobacteria and alpha- and beta-Proteobacteria (T2) showed increased metabolic potential with the possibility of uptake and processing of alternative carbon sources along with CO_2 fixation pathways. Metabolic potential was highest throughout the last three clusters, which were dominated by only one taxa each: Firmicutes (T3), a

second collection of alpha-Proteobacteria (T4) and finally gamma-Proteobacteria (T5; Figure 4.4).

To disentangle the metabolism of Firmicutes, alpha-Proteobacteria, gamma-Proteobacteria and that of the other taxa, statistical analysis of significant association was performed. Differences in carbon metabolism were seen throughout all taxa (Table 4.3, detailed table see Supplemental Table 7), most of these were associated with the catabolism of alternative carbon substrates to generate ATP. The substrates range from simple C2 to C6 molecules to more complex monosaccharides, oligosaccharides and nucleosides: most taxa were able to catabolize D-ribose (aldopentose). Pathways for the utilization of glycerol, glycerol-phosphate (both C-3 alcohols), glycerate (C-3 acid), lactate (C-3 alpha-hydroxy-acid), D-gluconate (C-6 acid), ketogluconates (C-6 acid), maltose (disaccharide), maltodextrin (oligosaccharide), deoxyribose and deoxynucleoside are significantly present in at least half of the taxonomic groups (Table 4.4). In order to metabolize this wide variety of substrates, some taxa employ alternative entries into the citric acid cycle. The glyoxylate bypass (which is also part of the serine-glyoxylate cycle) integrates short chain C-molecules into the citric acid cycle when glucose is not available (Lorenz et al. 2002; Zhao et al. 1996). 2-phosphoglycolate can be salvaged via the light-independent part of the photorespiration pathway, and ethylmalonyl-CoA is channeled via an alternative entry into the citric acid cycle (Ethylmalonyl-CoA pathway of C assimilation; Alber et al. 2006; Eisenhut et al. 2006). The Entner-Doudoroff Pathway, often found in organisms lacking enzymes needed for glycolysis, is present in half of the taxa.

Through this pathway, glucose is broken down less efficiently than during glycolysis resulting in 2 less ATP being generated (Entner and Doudoroff 1952). The diversity of carbon metabolism suggests that non-cyanobacterial diazotrophs are contributors to the cycling of the organic matter pool in the ocean and can break down a wide variety of organic compounds for energy generation.

Table 4.3: Carbon-related metabolic pathways significantly associated with major taxa of 132 diazotrophic reference genomes as annotated by SEED (Overbeek et al. 2005).

Taxa	Alternative carbon sources and alternative carbon metabolic pathways																Anaerobic respiration			Other		
	Ethylmalonyl-CoA pathway of C assimilation[1]	Mannitol Utilization[2]	Chitin and N-acetylglucosamine utilization[2]	Glyoxylate bypass[3]	Calvin-Benson cycle[4]	Photorespiration (oxidative C cycle)[5]	Lactate utilization[2]	Deoxyribose and Deoxynucleoside Catabolism[2]	Serine-glyoxylate cycle[6]	Glycerate metabolism[7]	D-gluconate and ketogluconates metabolism[2]	Entner-Doudoroff Pathway[8]	Maltose and Maltodextrin Utilization[2]	Glycerol and Glycerol phosphate Uptake and Utilization[2]	D-ribose utilization[2]	Carbon storage regulator[9]	Fermentations: Mixed acid[10]	Pyruvate:ferredoxin oxidoreductase[11]	Fermentations: Lactate[12]	Acetyl-CoA fermentation to Butyrate[13]	Methylglyoxal Metabolism[14]	Formaldehyde assimilation: Ribulose monophosphate pathway (RuMP)[15]	Methanogenesis[16]	Trehalose Biosynthesis[17]
Acidithiobacillia						+																		
Actinobacteria	+	+	+			+	+	+			+		+		+					+	+			
alpha-Proteobacteria				+	+	+	+		+	+	+		+		+		+			+	+			+
Archaea					+				+						+			+						
beta-Proteobacteria				+	+	+	+		+	+	+			+	+		+		+	+	+			
Chlorobi						+		+	+					+	+				+	+				
Chloroflexi	+			+		+	+	+			+		+		+					+	+			+
Cyanobacteria			+		+	+		+	+	+	+		+	+	+		+	+	+					
Deferribacteres									+	+					+	+	+	+	+					
delta-Proteobacteria							+	+		+			+	+	+	+	+	+	+	+	+			
epsilon-Proteobacteria																+		+						
Firmicutes								+				+	+	+		+		+		+	+			
gamma-Proteobacteria				+	+		+		+	+	+	+	+	+	+	+	+	+	+	+	+			
Verrucomicrobia	+		+			+	+	+		+	+	+	+	+	+	+	+	+	+		+			+
p-value[18]	0.0189	0.0412	0.0047	5.00E-04	0.01	1.00E-04	1.00E-04	0.0019	0.0025	0.0072	8.00E-04	3.00E-04	1.00E-04	8.00E-04	0.004	1.00E-04	0.0295	1.00E-04	0.001	1.00E-04	1.00E-04	0.0289	1.00E-04	8.00E-04

1. alternative C2 - molecule entering citric acid cycle
2. alternative C source •
3. citric acid cycle modification
4. most prominent of the autotrophic CO_2 fixation pathways
5. salvage pathway for 2-phosphoglycolate (light-independent)
6. citric acid cycle modification
7. C3-sugar acid utilization
8. alternative glycolysis
9. Regulation of carbon metabolism
10. anaerobic dissimilation of pyruvate to succinate, lactate, acetate, ethanol, formate, CO_2 and H_2
11. reduces pyruvate anaerobically
12. anaerobic dissimilation of pyruvate to lactate
13. Conversion from pyruvate to acetyl-CoA
14. detoxification
15. formaldehyde fixation and detoxification
16. anaerobic reduction of oxidized C compounds to methane
17. Disaccharide biosynthesis
18. subsystem functions were extracted from SEED (Overbeek et al. 2005). Their significance with particular taxonomic groups were tested in R using *indicspecies* (Caceres and Legendre 2009; R core team 2015)

There are several pathways indicating that most organisms analyzed possess the ability to respire anaerobically. In anaerobic respiration the final electron acceptor is not O_2, which results in less overall energy generation than in aerobic respiration due to a smaller electrochemical gradient between electron donors and acceptors. Final electron acceptors can be organic (halogenated organic compounds), inorganic (NO_3^-, SO_4^{3-} and CO_2, Fe, Mn, Co, S) or the products of fermentation. The fermentative processes of butyrate synthesis and the formation of mixed acids or lactate from pyruvate occur under anaerobic conditions (Pryde et al. 2002). Riemann et al. (2010) proposed that heterotrophic diazotrophs may fix N_2 in microaerobic niches on organic particles or in OMZs of the ocean. This is supported by our findings. The observed differences in anaerobic metabolism is related to the differences in enzymes that allow anaerobic regeneration of NAD^+ from pyruvate. These enzymes produce a variety of products (lactate or mixed-acid fermentation) and ultimately permit ATP synthesis without aerobic oxidative phosphorylation.

A part of carbon metabolism that has been of intensifying interest is the degradation of aromatic carbon compounds (Table 4.4). The potential for the degradation of benzoate (including toluene and salicylate-ester degradation), N-heterocyclic aromatic compounds, and gentisate (including xylenol and cresol), was found in the reference genomes of some diazotrophs. Alpha- and beta-Proteobacteria displayed all four pathways, whereas Actinobacteria, Cyanobacteria, Deferribacteres, gamma- and epsilon-Proteobacteria displayed only one or two. These pathways break down various toxic aromatic compounds

and will provide diazotrophs with additional carbon and nitrogen sources. These pathways are also extremely important in degrading crude oil and incomplete combustion products. In the marine environment, there is a focus on oil-degrading bacteria that could be used in bioremediation of oils spills (as for example extensively studied during the Deepwater Horizon oil spill in the Mexican Gulf; Dubinsky et al. 2013, Gutierrez et al. 2013). Oil-degrading diazotrophs would provide an exciting opportunity for bioengineering, with the ability to generate fixed nitrogen species in an environment that might be limited by nitrogen, thereby circumventing the standard practice of adding nitrate to stimulate bacterial growth during marine oil spill remediation (Swannell et al. 1996).

The last major metabolic variation in carbon metabolism is seen throughout most taxa and divides the group of non-cyanobacterial diazotrophs into heterotrophs and autotrophs (Supplemental Figure 11). We found two carbon fixation pathways present in the reference genomes: RuBisCO was present in 55 of the reference genomes among the Archaea, Cyanobacteria, Firmicutes, Chlorobi, alpha-, beta- and gamma-Proteobacteria and Verrucomicrobia (Supplemental Figure 11). Some Archaea and Acidithiobacillia showed the potential to incorporate the 1C compound formaldehyde via the ribulose-monophosphate-pathway (RuMP), which acts both in detoxification and as an assimilatory C pathway. There are several reference genomes with this pathway that are linked to the marine environment, either as the source of initial isolation or having been identified there since. These include an archaeon isolated from hydrothermal

vents (*Methanopyrus kandleri* AV19), four bacteria from the group Chlorobi found in aquatic environments and the Black Sea (*Chlorobaculum parvum* NCIB 8327, *Chloroherpeton thalassium* ATCC 35110, *Pelodictyon phaeoclathratiforme* BU-1 and Chlorobi*um phaeobacteroides* BS1), the alpha-Proteobacterium *Rhodospirillum centenum* SW , the gamma-Proteobacterium *Methylobacter luteus* IMV-B-3098 and the delta-Proteobacterium *Desulfovibrio vulgaris* str. Miyazaki F. With 59 of 132 reference genomes displaying the potential for non-photosynthetic inorganic carbon fixation, this may be a common mechanism to meet energy demands for N_2 fixation.

Table 4.4: Metabolic pathways related to aromatic compounds, nitrogen, phosphorus, sulfur- and iron that are significantly associated with 132 diazotrophic reference genomes as annotated by SEED (Overbeek et al. 2005).

Taxa	Aromatic				N		P	S			Fe				
	Benzoate degradation[1]	N-heterocyclic aromatic compound degradation[2]	Gentisate degradation[3]	Salicylate ester degradation	Denitrifying reductase gene clusters[4]	Nitrate and nitrite ammonification[5]	Alkylphosphonate utilization[6]	Utilization of glutathione as a sulphur source	Inorganic Sulfur Assimilation[7]	Sulfite reduction-associated complex DsrMKJOP and co clustering genes[8]	Salmochelin-mediated Iron Acquisition [9]	Siderophore assembly kit	Ferrous iron transporter EfeUOB, low pH induced	Heme, hemin uptake and utilization systems in Gram-Positives	Hemin transport system
Acidithiobacillia							+		+						
Actinobacteria			+	+		+	+	+	+		+	+	+		
alpha-Proteobacteria	+	+	+	+	+		+	+	+					+	+
Archaea															
beta-Proteobacteria	+	+	+	+	+	+	+		+						+
Chlorobi										+					
Chloroflexi														+	
Cyanobacteria				+		+			+					+	
Deferribacteres			+		+	+									+
delta-Proteobacteria					+	+									
epsilon-Proteobacteria				+	+	+			+						
Firmicutes										+					
gamma-Proteobacteria	+	+			+	+	+		+	+					+
Verrucomicrobia					+	+			+						+
p-value	0.0101	0.0241	0.0018	0.0122	0.0127	6.00E-04	1.00E-04	0.0401	1.00E-04	0.0352	0.0205	0.0163	0.0156	4.00E-04	0.0018

1. Several pathways of degradation of benzoate-like compounds including toluene
2. Degradation of azaarenes
3. Xylenols and cresols degradation
4. reducing nitrate to N_2 gas
5. Nitrate and nitrite reduction to ammonia
6. cleavage of C-P bonds
7. assimilatory sulfate reduction
8. enables sulfur oxidation in sulfur-oxidizing bacteria and sulfate and sulfide oxidation in sulfate- and sulfide-oxidizing bacteria and archaea
9. Salmochelins are glucosylated derivatives of enterobactin, which are secreted in response to iron deprivation

Although general phosphate metabolism pathways are present throughout all investigated taxa, the pathway of alkylphosphonate utilization is significantly present only in alpha-, beta- and gamma-Proteobacteria as well as Acidithiobacillia (Table 4.4). Natural and anthropogenic sources of alkylphosphonates can be utilized by microorganisms in areas limited in inorganic phosphate (McGrath et al. 2013). The ability to degrade phosphonates in nature is of particular selective advantage to diazotrophs since they are not growth-limited by nitrogen and has already been demonstrated for *Trichodesmium spp.* (Dyhrman et al. 2006). Degradation of alkylphosphonates has primarily been studied because these compounds are commonly used as herbicides and pesticides, and are also toxic to mammals (Singh and Allan 2006). These harmful compounds can be washed from farmland into rivers and ultimately into the ocean (Mercurio et al. 2014; Udiković-Kolić et al. 2012). Hence, the study of bacteria that can degrade these compounds is of interest to environmental and health studies.

Along with N_2 fixation, some diazotrophs participate in the marine nitrogen cycle in additional ways (Table 4.4). Perhaps most striking is the presence of genes for the denitrification pathway involved in the reduction of nitrate into nitrite or N_2 gas. This pathway is found throughout the Proteobacteria, Deferribacteres and Verrucomicrobia (analyzed here), showing these bacteria are capable of anaerobic respiration using nitrate as a substrate. This implies that the feedback mechanism between N_2 fixation and denitrification proposed by Deutsch et al (2007) may occur within the same organism. In general, oceanic matter displays

an N:P ratio of 16:1 (the Redfield Ratio; Redfield 1963). However, this ratio can deviate in regions of nitrogen loss (denitrification and anammox) or gain (N_2 fixation). From comparing N:P ratios along transported water masses, it was predicted that denitrification directly supports N_2 fixation, balancing nitrogen losses and gains in proximity (Deutsch et al. 2007). Four diazotrophic reference genomes among the alpha-, beta- and gamma-Proteobacteria have been found in the marine environment and show the potential to respire fixed nitrogen under anoxic conditions. All of them were isolated from low O_2 aquatic environments: *Magnetococcus sp*. (autotrophic alpha-Proteobacterium, isolated from the oxic-anoxic interface in the Pettaquamscutt Estuary, Frankel et al. 1997), *Rhodopseudomonas palustris* (heterotrophic alpha-Proteobacterium, isolated from suboxic Baltic Sea, Bentzon-Tilia et al. 2015), *Rhodospirillum centenum* (autotrophic alpha-Proteobacteria, isolated from an anoxic hot spring, but commonly identified in the marine environment; Favinger et al. 1989) and *Pseudomonas stutzeri* (heterotrophic gamma-Proteobacterium, isolated from suboxic Baltic Sea, Bentzon-Tilia et al. 2015).

Variations in the metabolism of sulfur compounds are related to variations in sulfur sources (inorganic sulfur assimilation and utilization of glutathione as a sulphur source) or the oxidation of sulfur containing compounds (Table 4.4; Sulfite reduction associated complex *DsrMKJOP* and co-clustering genes). The *DsrMKJOP* complex enables sulfur and sulfide oxidation in some bacteria and archaea (Dahl et al. 2005). These chemolithotrophic organisms use various reduced sulphur compounds as electron donors to generate energy (Pronk et al.

1990). Within the diazotrophic reference genomes, this ability is associated with Chlorobi, and alpha- and beta-Proteobacteria.

Since diazotrophs have a very high demand for iron, iron acquisition and metabolism were also investigated. There are a variety of mechanisms found throughout all taxa, including iron uptake using siderophores and the processing of heme and hemin and heme transporters (Table 4.4). The marine reference genomes of *Trichodesmium erythraeum IMS10* showed the potential to synthesize the siderophore enterobactin.

Ultimately, the high metabolic diversity of proteobacteria might support the finding that their *nifH* sequences are the most commonly isolated from the marine environment (Figure 4.2; Farnelid et al. 2011). Metabolic diversity, including the utilization of alternative carbon sources, detoxification, anaerobic respiration and iron scavenging, could assist in the adaptation to a variety of environments. It certainly shows that the metabolic function of diazotrophic Proteobacteria, as well as that of other non-cyanobacterial diazotrophs within marine microbial communities, is likely not limited to carrying out N_2 fixation.

4.6. Conclusions

The emerging view in the studies of oceanic diazotrophs points towards the fact that the non-cyanobacterial members should not be neglected. Studies from the 21st century have repeatedly shown that non-cyanobacterial diazotrophic are well distributed throughout all oceanic environments and that they make up most, if

not all, of the diazotrophic community below the photic zone and in low O_2 regimes. While the total *nifH* phylogenetic diversity is still unknown, recent advances in the application of high-throughput sequencing to the *nifH* gene has the potential to characterise the complete diversity captured by the commonly used *nifH* primers (Zehr et al. 2001). Though initial studies using this technique have already established the wide distribution of non-cyanobacterial diazotrophs, more research is needed to determine the complete distribution and diversity of these organisms (Bombar et al. 2016; Farnelid et al. 2011). A spatially- and temporally-resolved sampling regime should be considered to disentangle environmental factors driving the diversity and N_2 fixation rates of diazotrophic communities. Further novel diazotrophs likely exist in specific oceanic areas such as hydrothermal vents capable of supporting many specialized diazotrophic chemoauthotrophs and anaerobic organisms, or those that provide anaerobic conditions along with a nitrogen depleted environment compared to phosphorous, such as OMZs.

We are in need of cultivated diazotrophs from many of these regions to enable the in-depth exploration of their metabolism. The few marine non-cyanobacterial diazotrophs already cultivated have provided some insights into the non-cyanobacterial diazotrophic metabolism, including the optimal environmental conditions under which N_2 fixation is performed. However, until a larger number and variety of these organisms have been analyzed, true mechanisms remain speculation. Proposed ideas include a low N:P ratio or even a high C and Fe to

low N:P ratio, taking into account the heterotrophic need for organic substrates for growth and iron requirement of the nitrogenase.

To gain a primary understanding of the metabolism of non-cyanobacterial diazotrophs, metabolic pathways related to nutrient cycling were analyzed in reference genomes that most closely align with marine uncultured *nifH* sequences. This analysis revealed that diazotroph communities are adaptable to a variety of environmental conditions, including anaerobic water masses, and have diverse metabolic potential, including the utilization of a suite of organic carbon sources and various electron donors, as well as the degradation of toxic substances such as aromatic and phosphonate compounds. Pathways relevant to diazotrophy include: anaerobic respiration, as anaerobic conditions are favorable for N_2 fixation; the ability to perform denitrification, which may suggest an intra-organismal feedback between nitrogen gain and loss in the ocean; and the variety of mechanisms to salvage iron, which is essential for the nitrogenase enzyme. Diazotrophs capable of degrading toxic compounds could have bioremediation applications in nitrogen limited areas, such as remediation of oil spills in open ocean waters.

Altogether, this review has shown that non-cyanobacterial diazotrophs dominate the diazotrophic community in the aphotic, temperate and oxygen depleted ocean, where a widely varied nutrient metabolic potential allows them to inhabit diverse marine environments. Because N_2 fixation has been clearly demonstrated in oceanic regions lacking cyanobacteria where non-cyanobacterial diazotrophs have been identified, it seems logical to conclude that some of the identified non-

cyanobacterial diazotrophs are fixing nitrogen. Thus, it is highly likely that non-cyanobacterial diazotrophs significantly contribute to the total fixed N_2 fixation in the ocean.

4.7. Methods

To investigate the phylogeny of non-cyanobacterial diazotrophs, published *nifH* sequences greater than 300 base pairs in length were extracted from NCBI. After removal of duplicate and N-containing sequences, 3700 unique sequences of ca 360 base pair length remained (Table 4.1 and Table 4.2). The unique sequences' most closely related reference *nifH* sequences were obtained from the BLAST and SEED reference genome databases (Overbeek et al. 2005; Altschul et al. 1990). The closest reference genome of marine origin was selected wherever possible (Supplemental Table 7). In total, 132 reference genomes were identified. 16S rRNA gene and *nifH* sequences were extracted from the reference genomes for phylogenetic comparison to their *nifH* sequences (Figure 4.3). All *nifH* sequences were clustered at 96% nucleotide sequence similarity using CD-HIT, resulting in 655 operational taxonomic units (OTUs; Supplemental Table 9; Li and Godzik 2006). 16S rRNA genes were aligned based on their nucleotide sequence and *nifH* gene sequences were aligned based on their protein sequence using MAFFT v. 7 (Yamada et al. 2016; Katoh et al. 2002). Protein alignments were converted to nucleotide sequences (PAL2NAL; Suyama et al. 2006) and ambiguous sequence alignment regions were removed using Gblocks (Castresana 2000) reducing alignment positions from 444 to 318 for *nifH* and

from 2343 to 1314 for 16S rRNA before inferring maximum likelihood trees based on the GTR-GAMMA model in RAxML using default parameters (Stamakatis 2014). A protein maximum likelihood tree was constructed from clustered sequences based on the WAG-GAMMA model in RAxML using default parameters (Stamakatis 2014). Bootstrap values were calculated with the rapid bootstrapping algorithm from 100 maximum likelihood tree replicates. The trees were displayed with branch lengths showing the number of substitutions per site in iTOL (Letunic and Borg 2016).

To investigate the metabolic potential of non-cyanobacterial diazotrophs, reference genomes were annotated using FROMP and SEED (Desai et al. 2013; Overbeek et al. 2005). The Statistical program PRIMER was used to identify metabolic differences between taxonomic groups. A matrix was generated showing the presence and absence of 167 specific nutrient cycles (carbon, aromatic compounds, nitrogen, phosphate, sulfur and iron) in each reference genome as annotated using FROMP (Desai et al. 2013). An ANOSIM test was performed to establish metabolic differences between taxonomic groups and significant pairwise differences are recorded in Supplemental Table 6. Pathways that were not present in at least three reference genomes were removed from further analysis. A rough association of the remaining 142 specific pathways with the 132 reference genomes was established by clustering analysis in the R package *gplots* and displayed as a heatmap using *ggplot2* (R core team 2015; Wickham and Chang 2015; Warnes et al. 2013). Then, the specific association of metabolic pathways with taxa was determined statistically in R using the

multilevel pattern analysis *indicspecies* (R core team 2015; Caceres and Legendre 2009). Reference genomes were assigned to one of the following taxonomic groups: Archaea, Acidithiobacillia, Actinobacteria, Chlorobi, Chloroflexi, Cyanobacteria, Deferribacteres, Firmicutes, alpha-Proteobacteria, beta-Proteobacteria, gamma-Proteobacteria, delta-Proteobacteria, epsilon-Proteobacteria and Verrucomicrobia. Summaries of metabolic pathways with significant association to specific taxa (p-value < 0.05) are displayed in Table 4.3 (carbon metabolism) and Table 4.4 (aromatic compound, nitrogen, phosphate, sulfur and iron metabolism). Detailed tables are found in Supplemental Table 7 and Supplemental Table 8.

4.8. Acknowledgments

Jenni-Marie Ratten's work was supported by the NSERC CREATE Transatlantic Ocean System Science and Technology (TOSST) grant, the Stiftung für Kanadastudien (SKS), the Deutscher Akademischer Austausch Dienst (DAAD) and Julie LaRoche's NSERC Discovery research grant. We would like to thank Jennifer Tolman, Sallie Lau, Jackie Zorz and Hannah Blanchard for proof reading the document.

4.9. References

Abell, J., Emerson, S. and Renaud, P., 2000. Distributions of TOP, TON and TOC in the North Pacific subtropical gyre: Implications for nutrient supply in the surface ocean and remineralization in the upper thermocline. Journal of Marine Research, 58(2), pp.203-222.

Alber, B.E., Spanheimer, R., Ebenau-Jehle, C. and Fuchs, G., 2006. Study of an alternate glyoxylate cycle for acetate assimilation by Rhodobacter sphaeroides. Molecular microbiology, 61(2), pp.297-309.

Altabet, M.A., 2006. Constraints on oceanic N balance/imbalance from sedimentary 15 N records. Biogeosciences Discussions, 3(4), pp.1121-1155.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

Andersson, B., Sundbäck, K., Hellman, M., Hallin, S. and Alsterberg, C., 2014. Nitrogen fixation in shallow-water sediments: Spatial distribution and controlling factors. Limnology and Oceanography, 59(6), pp.1932-1944.

Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stöckel, J., Liberton, M., Min, H., Sherman, L.A. and Pakrasi, H.B., 2011. Novel metabolic attributes of the genus Cyanothece, comprising a group of unicellular nitrogen-fixing cyanobacteria. MBio, 2(5), pp.e00214-11.

Béja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L. and Spudich, E.N., 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science, 289(5486), pp.1902-1906.

Benavides, M., Bonnet, S., Hernández, N., Martínez-Pérez, A.M., Nieto-Cid, M., Álvarez-Salgado, X.A., Baños, I., Montero, M.F., Mazuecos, I.P., Gasol, J.M. and Osterholz, H., 2016. Basin-wide N_2 fixation in the deep waters of the Mediterranean Sea. Global Biogeochemical Cycles, 30(6), pp.952-961.

Bentzon-Tilia, M., Farnelid, H., Jürgens, K. and Riemann, L., 2014. Cultivation and isolation of N_2-fixing bacteria from suboxic waters in the Baltic Sea. FEMS microbiology ecology, 88(2), pp.358-371.

Bentzon-Tilia, M., Traving, S.J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J.L., Markager, S. and Riemann, L., 2015. Significant N_2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. The ISME journal, 9(2), pp.273-285.

Berelson, W.M., 1991. The flushing of two deep-sea basins, southern California borderland. Limnology and Oceanography, 36(6), pp.1150-1166.

Berman-Frank, I., Cullen, J.T., Shaked, Y., Sherrell, R.M. and Falkowski, P.G., 2001. Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. Limnology and Oceanography, 46(6), pp.1249-1260.

Berman-Frank, I., Lundgren, P. and Falkowski, P., 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Research in Microbiology, 154(3), pp.157-164.

Berman-Frank, I., Quigg, A., Finkel, Z.V., Irwin, A.J. and Haramaty, L., 2007. Nitrogen-fixation strategies and Fe requirements in cyanobacteria. Limnology and Oceanography, 52(5), pp.2260-2269.

Bertics, V.J., Löscher, C.R., Salonen, I., Dale, A.W., Gier, J., Schmitz, R.A. and Treude, T., 2013. Occurrence of benthic microbial nitrogen fixation coupled to sulfate reduction in the seasonally hypoxic Eckernförde Bay, Baltic Sea. Biogeosciences, 10(3), pp.1243-1258.

Bethoux, J.P., Morin, P., Madec, C. and Gentili, B., 1992. Phosphorus and nitrogen behaviour in the Mediterranean Sea. Deep Sea Research Part A. Oceanographic Research Papers, 39(9), pp.1641-1654.

Bible, A.N., Khalsa-Moyers, G.K., Mukherjee, T., Green, C.S., Mishra, P., Purcell, A., Aksenova, A., Hurst, G.B. and Alexandre, G., 2015. Metabolic adaptations of Azospirillum brasilense to oxygen stress by cell-to-cell clumping and flocculation. Applied and environmental microbiology, 81(24), pp.8346-8357.

Bird, C., Martinez, J.M., O'Donnell, A.G. and Wyman, M., 2005. Spatial distribution and transcriptional activity of an uncultured clade of planktonic diazotrophic γ-Proteobacteria in the Arabian Sea. Applied and environmental microbiology, 71(4), pp.2079-2085.

Bird, C. and Wyman, M., 2013. Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea. FEMS microbiology ecology, 84(1), pp.189-200.

Bombar, D., Moisander, P.H., Dippner, J.W., Foster, R.A., Voss, M., Karfeld, B. and Zehr, J.P., 2011. Distribution of diazotrophic microorganisms and nifH gene expression in the Mekong River plume during intermonsoon. Marine Ecology Progress Series, 424, pp.39-52.

Bombar, D., Paerl, R.W. and Riemann, L., 2016. Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. Trends in microbiology, 24(11), pp.916-927.

Bombar, D., Turk-Kubo, K.A., Robidart, J., Carter, B.J. and Zehr, J.P., 2013. Non-cyanobacterial nifH phylotypes in the North Pacific Subtropical Gyre detected by flow-cytometry cell sorting. Environmental microbiology reports, 5(5), pp.705-715.

Bonnet, S., Guieu, C., Bruyant, F., Prášil, O., Van Wambeke, F., Raimbault, P., Moutin, T., Grob, C., Gorbunov, M.Y., Zehr, J.P. and Masquelier, S.M., 2008. Nutrient limitation of primary productivity in the Southeast Pacific (BIOSOPE cruise). Biogeosciences, 5(1), pp.215-225.

Bonnet, S., Grosso, O. and Moutin, T., 2011. Planktonic N_2 fixation in the Mediterranean Sea: a major biogeochemical process during the stratified period. Biogeosciences Discussions, 8(1), pp.1197-1225.

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K.A., Moutin, T., Hamersley, R.M., Grosso, O., Zehr, J.P. and Capone, D.G., 2013. Aphotic N 2 fixation in the eastern tropical South Pacific Ocean. Plos one, 8(12), p.e81265.

Capone, D.G., Burns, J.A., Montoya, J.P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A.F. and Carpenter, E.J., 2005. Nitrogen fixation by

Trichodesmium spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. Global Biogeochemical Cycles, 19(2).

Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B. and Carpenter, E.J., 1997. *Trichodesmium*, a globally significant marine cyanobacterium. Science, 276(5316), pp.1221-1229.

Carpenter, E.J., Montoya, J.P., Burns, J., Mulholland, M.R., Subramaniam, A. and Capone, D.G., 1999. Extensive bloom of a N_2-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), pp.540-552.

Cáceres, M.D. and Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology, 90(12), pp.3566-3574.

Cheung, S., Xia, X., Guo, C. and Liu, H., 2016. Diazotroph community structure in the deep oxygen minimum zone of the Costa Rica Dome. Journal of plankton research, 38(2), pp.380-391.

Church, M.J., Bjorkman, K.M., Karl, D.M., Saito, M.A. and Zehr, J.P., 2008. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. Limnology and Oceanography, 53(1), p.63.

Church, M.J., Jenkins, B.D., Karl, D.M. and Zehr, J.P., 2005. Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. Aquatic Microbial Ecology, 38(1), pp.3-14.

Church, M.J., Mahaffey, C., Letelier, R.M., Lukas, R., Zehr, J.P. and Karl, D.M., 2009. Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. Global Biogeochemical Cycles, 23(2).

Clarke, K.R. and Gorley, R.N., 2006. PRIMER V6: user manual-tutorial. Plymouth Marine Laboratory.

Codispoti, L.A., 2007. An oceanic fixed nitrogen sink exceeding 400 Tg N a-1 vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences, 4(2), pp.233-253.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W. and Yoshinari, T., 2001. The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene?. Scientia Marina, 65(S2), pp.85-105.

Conley, D.J., Humborg, C., Rahm, L., Savchuk, O.P. and Wulff, F., 2002. Hypoxia in the Baltic Sea and basin-scale changes in phosphorus biogeochemistry. Environmental science & technology, 36(24), pp.5315-5320.

Cook, P.L., Evrard, V. and Woodland, R.J., 2015. Factors controlling nitrogen fixation in temperate seagrass beds. Marine Ecology Progress Series, 525, pp.41-51.

Dahl, C., Engels, S., Pott-Sperling, A.S., Schulte, A., Sander, J., Lübbe, Y., Deuster, O. and Brune, D.C., 2005. Novel genes of the dsr gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium Allochromatium vinosum. Journal of bacteriology, 187(4), pp.1392-1404.

Daneri, G., Dellarossa, V., Quiñones, R., Jacob, B., Montero, P. and Ulloa, O., 2000. Primary production and community respiration in the Humboldt Current System off Chile and associated oceanic areas. Marine Ecology Progress Series, 197, pp.41-49.

Dekaezemacker, J., Bonnet, S., Grosso, O., Moutin, T., Bressac, M. and Capone, D.G., 2013. Evidence of active dinitrogen fixation in surface waters of the eastern tropical South Pacific during El Niño and La Niña events and evaluation of its potential nutrient controls. Global Biogeochemical Cycles, 27(3), pp.768-779.

Dekas, A.E., Poretsky, R.S. and Orphan, V.J., 2009. Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. Science, 326(5951), pp.422-426.

Desai, M.S. and Brune, A., 2012. Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. The ISME journal, 6(7), pp.1302-1313.

Desai, D.K., Schunck, H., Löser, J.W. and LaRoche, J., 2013. Fragment recruitment on metabolic pathways: comparative metabolic profiling of metagenomes and metatranscriptomes. Bioinformatics, 29(6), pp.790-791.

Díez, B., Bergman, B., Pedrós-Alió, C., Antó, M. and Snoeijs, P., 2012. High cyanobacterial nifH gene diversity in Arctic seawater and sea ice brine. Environmental microbiology reports, 4(3), pp.360-366.

Dingler, C. and Oelze, J., 1987. Superoxide dismutase and catalase in Azotobacter vinelandii grown in continuous culture at different dissolved oxygen concentrations. Archives of microbiology, 147(3), pp.291-294.

Dingler, C., Kuhla, J., Wassink, H. and Oelze, J., 1988. Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations. Journal of bacteriology, 170(5), pp.2148-2152.

Dixon, R. and Kahn, D., 2004. Genetic regulation of biological nitrogen fixation. Nature Reviews Microbiology, 2(8), pp.621-631.

Dubinsky, E.A., Conrad, M.E., Chakraborty, R., Bill, M., Borglin, S.E., Hollibaugh, J.T., Mason, O.U., M. Piceno, Y., Reid, F.C., Stringfellow, W.T. and Tom, L.M., 2013. Succession of hydrocarbon-degrading bacteria in the aftermath of the Deepwater Horizon oil spill in the Gulf of Mexico. Environmental science & technology, 47(19), pp.10860-10867.

Dyhrman, S.T., Chappell, P.D., Haley, S.T., Moffett, J.W., Orchard, E.D., Waterbury, J.B. and Webb, E.A., 2006. Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. Nature, 439(7072), pp.68-71.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), pp.1792-1797.

Eisenhut, M., Kahlon, S., Hasse, D., Ewald, R., Lieman-Hurwitz, J., Ogawa, T., Ruth, W., Bauwe, H., Kaplan, A. and Hagemann, M., 2006. The plant-like C2 glycolate cycle and the bacterial-like glycerate pathway cooperate in phosphoglycolate metabolism in cyanobacteria. Plant Physiology, 142(1), pp.333-342.

Entner, N. and Doudoroff, M., 1952. Glucose and gluconic acid oxidation of Pseudomonas saccharophila. Journal of Biological Chemistry, 196(2), pp.853-862.

Falcón, L.I., Carpenter, E.J., Cipriano, F., Bergman, B. and Capone, D.G., 2004. N_2 fixation by unicellular bacterioplankton from the Atlantic and Pacific Oceans: Phylogeny and in situ rates. Applied and Environmental Microbiology, 70(2), pp.765-770.

Falkowski, P.G., 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO 2 in the ocean. Nature, 387(6630), pp.272-275.

Falkowski, P.G., 1983. Enzymology of nitrogen assimilation. Nitrogen in the marine environment, pp.839-868.

Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å. and Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PLoS One, 6(4), p.e19223.

Farnelid, H., Bentzon-Tilia, M., Andersson, A.F., Bertilsson, S., Jost, G., Labrenz, M., Jürgens, K. and Riemann, L., 2013. Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. The ISME journal, 7(7), pp.1413-1423.

Farnelid, H., Harder, J., Bentzon-Tilia, M. and Riemann, L., 2014. Isolation of heterotrophic diazotrophic bacteria from estuarine surface waters. Environmental microbiology, 16(10), pp.3072-3082.

Farnelid, H., Öberg, T. and Riemann, L., 2009. Identity and dynamics of putative N_2-fixing picoplankton in the Baltic Sea proper suggest complex patterns of regulation. Environmental microbiology reports, 1(2), pp.145-154.

Farnelid, H., Tarangkoon, W., Hansen, G., Hansen, P.J. and Riemann, L., 2010. Putative N_2-fixing heterotrophic bacteria associated with dinoflagellate–Cyanobacteria consortia in the low-nitrogen Indian Ocean. Aquatic Microbial Ecology, 61(2), pp.105-117.

Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å. and Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PLoS One, 6(4), p.e19223.

Favinger, J., Stadtwald, R. and Gest, H., 1989. Rhodospirillum centenum, sp. nov., a thermotolerant cyst-forming anoxygenic photosynthetic bacterium. Antonie van Leeuwenhoek, 55(3), pp.291-296.

Fernandez, C., Farías, L. and Ulloa, O., 2011. Nitrogen fixation in denitrified marine waters. PLoS One, 6(6), p.e20539.

Fernandez, C., González, M.L., Muñoz, C., Molina, V. and Farias, L., 2015. Temporal and spatial variability of biological nitrogen fixation off the upwelling system of central Chile (35–38.5° S). Journal of Geophysical Research: Oceans, 120(5), pp.3330-3349.

Foster, R.A., Paytan, A. and Zehr, J.P., 2009. Seasonality of N_2 fixation and nifH gene diversity in the Gulf of Aqaba (Red Sea). Limnology and Oceanography, 54(1), pp.219-233.

Frankel, R.B., Bazylinski, D.A., Johnson, M.S. and Taylor, B.L., 1997. Magneto-aerotaxis in marine coccoid bacteria. Biophysical journal, 73(2), p.994.

Fredriksson, C. and Bergman, B., 1997. Ultrastructural characterisation of cells specialised for nitrogen fixation in a non-heterocystous cyanobacterium, *Trichodesmium* spp. Protoplasma, 197(1-2), pp.76-85.

Fuller, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F. and Scanlan, D.J., 2005. Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. Limnology and Oceanography, 50(1), pp.363-375.

Giuffrè, A., Borisov, V.B., Arese, M., Sarti, P. and Forte, E., 2014. Cytochrome bd oxidase and bacterial tolerance to oxidative and nitrosative stress. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1837(7), pp.1178-1187.

Goto, M., Ando, S., Hachisuka, Y. and Yoneyama, T., 2005. Contamination of diverse nifH and nifH-like DNA into commercial PCR primers. FEMS microbiology letters, 246(1), pp.33-38.

Großkopf, T. and LaRoche, J., 2012. Direct and indirect costs of N_2 fixation in Crocosphaera watsonii WH8501 and possible implications for the nitrogen cycle. Frontiers in microbiology, 3, p.236.

Großkopf, T., Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M.M., Lavik, G., Schmitz, R.A., Wallace, D.W. and LaRoche, J., 2012. Doubling of marine N_2-fixation rates based on direct measurements. Nature, 488(7411), pp.361-364.

Gruber, N., 2008. The marine nitrogen cycle: overview and challenges. Nitrogen in the marine environment, pp.1-50.

Gutierrez, T., Singleton, D.R., Berry, D., Yang, T., Aitken, M.D. and Teske, A., 2013. Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. The ISME journal, 7(11), pp.2091-2104.

Halm, H., Lam, P., Ferdelman, T.G., Lavik, G., Dittmar, T., LaRoche, J., D'Hondt, S. and Kuypers, M.M., 2012. Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. The ISME journal, 6(6), pp.1238-1249.

Hamersley, M.R., Lavik, G., Woebken, D., Rattray, J.E., Lam, P., Hopmans, E.C., Damsté, J.S.S., Krüger, S., Graco, M., Gutiérrez, D. and Kuypers, M.M., 2007. Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. Limnology and Oceanography, 52(3), pp.923-933.

Hamersley, M.R., Turk, K.A., Leinweber, A., Gruber, N., Zehr, J.P., Gunderson, T. and Capone, D.G., 2011. Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. Aquatic Microbial Ecology, 63(2), pp.193-205.

Haselkorn, R., 2007. Heterocyst differentiation and nitrogen fixation in cyanobacteria. In Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations (pp. 233-255). Springer Netherlands.

Hashimoto, R., Yoshida, T., Kuno, S., Nishikawa, T. and Sako, Y., 2012. The first assessment of cyanobacterial and diazotrophic diversities in the Japan Sea. Fisheries Science, 78(6), pp.1293-1300.

Hewson, I., Moisander, P.H., Achilles, K.M., Carlson, C.A., Jenkins, B.D., Mondragon, E.A., Morrison, A.E. and Zehr, J.P., 2007. Characteristics of diazotrophs in surface to abyssopelagic waters of the Sargasso Sea. Aquatic microbial ecology, 46(1), pp.15-30.

Hilton, J.A., Satinsky, B.M., Doherty, M., Zielinski, B. and Zehr, J.P., 2015. Metatranscriptomics of N_2-fixing cyanobacteria in the Amazon River plume. The ISME journal, 9(7), pp.1557-1569.

Holl, C.M., Waite, A.M., Pesant, S., Thompson, P.A. and Montoya, J.P., 2007. Unicellular diazotrophy as a source of nitrogen to Leeuwin Current coastal eddies. Deep Sea Research Part II: Topical Studies in Oceanography, 54(8), pp.1045-1054.

Ingall, E. and Jahnke, R., 1994. Evidence for enhanced phosphorus regeneration from marine sediments overlain by oxygen depleted waters. Geochimica et Cosmochimica Acta, 58(11), pp.2571-2575.

Inomura, K., Bragg, J. and Follows, M.J., 2017. A quantitative analysis of the direct and indirect costs of nitrogen fixation: a model based on Azotobacter vinelandii. The ISME Journal, 11(1), pp.166-175.

Izquierdo, J.A. and Nüsslein, K., 2006. Distribution of extensive nifH gene diversity across physical soil microenvironments. Microbial Ecology, 51(4), pp.441-452.

Jahson, S., Rai, A.N. and Bergman, B., 1995. Intracellular cyanobiont Richelia intracellularis: ultrastructure and immuno-localisation of phycoerythrin, nitrogenase, Rubisco and glutamine synthetase. Marine Biology, 124(1), pp.1-8.

Janssen, F., Schrum, C. and Backhaus, J.O., 1999. A climatological data set of temperature and salinity for the Baltic Sea and the North Sea. Deutsche Hydrografische Zeitschrift, 51(9), p.5.

Jayakumar, A., Al-Rshaidat, M.M., Ward, B.B. and Mulholland, M.R., 2012. Diversity, distribution, and expression of diazotroph nifH genes in oxygen-deficient waters of the Arabian Sea. FEMS microbiology ecology, 82(3), pp.597-606.

Jickells, T.D., An, Z.S., Andersen, K.K., Baker, A.R., Bergametti, G., Brooks, N., Cao, J.J., Boyd, P.W., Duce, R.A., Hunter, K.A. and Kawahata, H., 2005. Global iron connections between desert dust, ocean biogeochemistry, and climate. science, 308(5718), pp.67-71.

Johnston, A.W., Li, Y. and Ogilvie, L., 2005. Metagenomic marine nitrogen fixation–feast or famine?. Trends in microbiology, 13(9), pp.416-420.

Juenemann, S., Butterworth, P.J. and Wrigglesworth, J.M., 1995. A suggested mechanism for the catalytic cycle of cytochrome bd terminal oxidase based on kinetic analysis. Biochemistry, 34(45), pp.14861-14867.

Jungblut, A.D., Lovejoy, C. and Vincent, W.F., 2010. Global distribution of cyanobacterial ecotypes in the cold biosphere. The ISME Journal, 4(2), pp.191-202.

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D. and Stal, L., 2002. N_2 fixation in the world's

oceans. In The Nitrogen Cycle at Regional to Global Scales (pp. 47-98). Springer Netherlands.

Knapp, A.N., 2015. The sensitivity of marine N 2 fixation to dissolved inorganic nitrogen. The microbial nitrogen cycle, p.90.

Kolonay, J.F., Moshiri, F., Gennis, R.B., Kaysser, T.M. and Maier, R.J., 1994. Purification and characterization of the cytochrome bd complex from Azotobacter vinelandii: comparison to the complex from Escherichia coli. Journal of bacteriology, 176(13), pp.4177-4181.

Kong, L., Jing, H., Kataoka, T., Sun, J. and Liu, H., 2011. Phylogenetic diversity and spatio-temporal distribution of nitrogenase genes (nifH) in the northern South China Sea. Aquatic Microbial Ecology, 65(1), pp.15-27.

Krom, M.D., Kress, N., Brenner, S. and Gordon, L.I., 1991. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. Limnology and Oceanography, 36(3), pp.424-432.

Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B.M., Amann, R.I., Kuypers, M.M. and Foster, R.A., 2013. In situ identification and N 2 and C fixation rates of uncultivated cyanobacteria populations. Systematic and applied microbiology, 36(4), pp.259-271.

Krupke, A., Mohr, W., LaRoche, J., Fuchs, B.M., Amann, R.I. and Kuypers, M.M., 2015. The effect of nutrients on carbon and nitrogen fixation by the UCYN-A–haptophyte symbiosis. The ISME journal, 9(7), pp.1635-1647.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution, p.msw054.

Kustka, A.B., Sanudo-Wilhelmy, S.A., Carpenter, E.J., Capone, D., Burns, J. and Sunda, W.G., 2003. Iron requirements for N_2-and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations. Limnology and Oceanography, 48(5), pp.1869-1884.

Labiosa, R.G., Arrigo, K.R., Genin, A., Monismith, S.G. and van Dijken, G., 2003. The interplay between upwelling and deep convective mixing in determining the seasonal phytoplankton dynamics in the Gulf of Aqaba: Evidence from SeaWiFS and MODIS. Limnology and oceanography, 48(6), pp.2355-2368.

Langlois, R.J., Hümmer, D. and LaRoche, J., 2008. Abundances and distributions of the dominant nifH phylotypes in the Northern Atlantic Ocean. Applied and environmental microbiology, 74(6), pp.1922-1931.

Langlois, R.J., LaRoche, J. and Raab, P.A., 2005. Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. Applied and Environmental Microbiology, 71(12), pp.7910-7919.

Langlois, R., Großkopf, T., Mills, M., Takeda, S. and LaRoche, J., 2015. Widespread distribution and expression of gamma A (UMB), an uncultured, diazotrophic, γ-proteobacterial nifH phylotype. PloS one, 10(6), p.e0128912.

Lema, K.A., Willis, B.L. and Bourne, D.G., 2012. Corals form characteristic associations with symbiotic nitrogen-fixing bacteria. Applied and environmental microbiology, 78(9), pp.3136-3144.

Letunic, I. and Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, p.gkw290.

Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), pp.1658-1659.

Loescher, C.R., Großkopf, T., Desai, F.D., Gill, D., Schunck, H., Croot, P.L., Schlosser, C., Neulinger, S.C., Pinnow, N., Lavik, G. and Kuypers, M.M., 2014. Facets of diazotrophy in the oxygen minimum zone waters off Peru. The ISME journal, 8(11), pp.2180-2192.

Lorenz, M.C. and Fink, G.R., 2002. Life and death in a macrophage: role of the glyoxylate cycle in virulence. Eukaryotic cell, 1(5), pp.657-662.

Luo, Y.W., Doney, S.C., Anderson, L.A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K.H., Böttjer, D., Capone, D.G. and Carpenter, E.J., 2012. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. Earth System Science Data, 4(1), pp.47-73.

Madigan, M.T., 1995. Microbiology of nitrogen fixation by anoxygenic photosynthetic bacteria. In Anoxygenic photosynthetic bacteria (pp. 915-928). Springer Netherlands.

Mahaffey, C., Michaels, A.F. and Capone, D.G., 2005. The conundrum of marine N_2 fixation. American Journal of Science, 305(6-8), pp.546-595.

Man-Aharonovich, D., Kress, N., Zeev, E.B., Berman-Frank, I. and Béjà, O., 2007. Molecular ecology of nifH genes and transcripts in the eastern Mediterranean Sea. Environmental microbiology, 9(9), pp.2354-2363.

Manasrah, R., Zibdah, M., Al-Ougaily, F., Yusuf, N. and Al-Najjar, T., 2007. Seasonal changes of water properties and current in the northernmost Gulf of Aqaba, Red Sea. Ocean Science Journal, 42(2), pp.103-116.

Marchal, K. and Vanderleyden, J., 2000. The" oxygen paradox" of N_2-fixing bacteria. Biology and fertility of soils, 30(5-6), pp.363-373.

Martinez-Perez, C., Mohr, W., Löscher, C.R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B.M., Lavik, G., Schmitz, R.A. and LaRoche, J., 2016. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. Nature Microbiology, 1, p.16163.

Marty, D.G., 1993. Methanogenic bacteria in seawater. Limnology and oceanography, 38(2), pp.452-456.

McGrath, J.W., Chin, J.P. and Quinn, J.P., 2013. Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. Nature Reviews Microbiology, 11(6), pp.412-419.

Mehta, M.P., Butterfield, D.A. and Baross, J.A., 2003. Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge. Applied and Environmental Microbiology, 69(2), pp.960-970.

Mercurio, P., Flores, F., Mueller, J.F., Carter, S. and Negri, A.P., 2014. Glyphosate persistence in seawater. Marine pollution bulletin, 85(2), pp.385-390.

Moffett, J.W., Goepfert, T.J. and Naqvi, S.W.A., 2007. Reduced iron associated with secondary nitrite maxima in the Arabian Sea. Deep Sea Research Part I: Oceanographic Research Papers, 54(8), pp.1341-1349.

Mohr, W., Intermaggio, M.P. and LaRoche, J., 2010. Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium Crocosphaera watsonii WH8501. Environmental microbiology, 12(2), pp.412-421.

Moisander, P.H., Beinart, R.A., Voss, M. and Zehr, J.P., 2008. Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. The ISME journal, 2(9), pp.954-967.

Moisander, P.H., Serros, T., Paerl, R.W., Beinart, R.A. and Zehr, J.P., 2014. Gammaproteobacterial diazotrophs and nifH gene expression in surface waters of the South Pacific Ocean. The ISME journal, 8(10), pp.1962-1973.

Moisander, P.H., Beinart, R.A., Hewson, I., White, A.E., Johnson, K.S., Carlson, C.A., Montoya, J.P. and Zehr, J.P., 2010. Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. Science, 327(5972), pp.1512-1514.

Moore, C.M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., Galbraith, E.D., Geider, R.J., Guieu, C., Jaccard, S.L. and Jickells, T.D., 2013. Processes and patterns of oceanic nutrient limitation. Nature Geoscience, 6(9), pp.701-710.

Moshiri, F., Crouse, B.R., Johnson, M.K. and Maier, R.J., 1995. The" nitrogenase-protective" FeSII protein of Azotobacter vinelandii: overexpression, characterization, and crystallization. Biochemistry, 34(40), pp.12973-12982.

Moutin, T., Karl, D.M., Duhamel, S., Rimmelin, P., Raimbault, P., Van Mooy, B.A. and Claustre, H., 2007. Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean. Biogeosciences Discussions, 4(4), pp.2407-2440.

Mulholland, M.R., Bernhardt, P.W., Blanco-Garcia, J.L., Mannino, A., Hyde, K., Mondragon, E., Turk, K., Moisander, P.H. and Zehr, J.P., 2012. Rates of N_2 fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank.

Naqvi, S.W.A., 2008. The Indian Ocean.

Needoba, J.A., Foster, R.A., Sakamoto, C., Zehr, J.P. and Johnson, K.S., 2007. Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. Limnology and Oceanography, 52(4), pp.1317-1327.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R. and Fonstein, M., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research, 33(17), pp.5691-5702.

Paerl, H.W. and Prufert, L.E., 1987. Oxygen-poor microzones as potential sites of microbial N_2 fixation in nitrogen-depleted aerobic marine waters. Applied and Environmental Microbiology, 53(5), pp.1078-1087.

Pantoja, S., Repeta, D.J., Sachs, J.P. and Sigman, D.M., 2002. Stable isotope constraints on the nitrogen cycle of the Mediterranean Sea water column. Deep Sea Research Part I: Oceanographic Research Papers, 49(9), pp.1609-1621.

Paulus, A., Rossius, S.G.H., Dijk, M. and de Vries, S., 2012. Oxoferryl-porphyrin radical catalytic intermediate in cytochrome bd oxidases protects cells from formation of reactive oxygen species. Journal of Biological Chemistry, 287(12), pp.8830-8838.

Ploug, H., Kühl, M., Buchholz-Cleven, B. and Jørgensen, B.B., 1997. Anoxic aggregates-an ephemeral phenomenon in the pelagic environment?. Aquatic Microbial Ecology, 13(3), pp.285-294.

Poole, R.K. and Hill, S., 1997. Respiratory protection of nitrogenase activity in Azotobacter vinelandii—roles of the terminal oxidases. Bioscience reports, 17(3), pp.303-317.

Pronk, J.T., Meulenberg, R., Hazeu, W., Bos, P. and Kuenen, J.G., 1990. Oxidation of reduced inorganic sulphur compounds by acidophilic thiobacilli. FEMS Microbiology letters, 75(2-3), pp.293-306.

Pryde, S.E., Duncan, S.H., Hold, G.L., Stewart, C.S. and Flint, H.J., 2002. The microbiology of butyrate formation in the human colon. FEMS microbiology letters, 217(2), pp.133-139.

Rahav, E., Bar-Zeev, E., Ohayion, S., Elifantz, H., Belkin, N., Herut, B., Mulholland, M.R. and Berman-Frank, I.R., 2013c. N_2 fixation in aphotic oxygenated marine environments. Frontiers in microbiology, 4, p.227.

Rahav, E., Herut, B., Mulholland, M.R., Belkin, N., Elifantz, H. and Berman-Frank, I., 2015. Heterotrophic and autotrophic contribution to N_2 fixation in the Gulf of Aqaba. Marine Ecology Progress Series, 522, pp.67-77.

Rahav, E., Herut, B., Mulholland, M.R., Voß, B., Stazic, D., Steglich, C., Hess, W.R. and Berman-Frank, I., 2013b. Contribution of N_2 fixation to bacterial and primary productivity in the Gulf of Aqaba (Red Sea). Biogeosciences Discussions, 10(6), pp.10327-10361.

Rahav, E., Herut, B., Stambler, N., Bar-Zeev, E., Mulholland, M.R. and Berman-Frank, I., 2013a. Uncoupling between N_2 fixation and primary productivity in the eastern Mediterranean Sea. Journal of Geophysical Research: Biogeosciences, 118(1), pp.195-202.

Raimbault, P. and Garcia, N., 2007. Carbon and nitrogen uptake in the South Pacific Ocean: evidence for efficient dinitrogen fixation and regenerated production leading to large accumulation of dissolved organic matter in nitrogen-depleted waters. Biogeosciences Discussions, 4(5), pp.3531-3579.

Ratten, J.M., LaRoche, J., Desai, D.K., Shelley, R.U., Landing, W.M., Boyle, E., Cutter, G.A. and Langlois, R.J., 2015. Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. Deep Sea Research Part II: Topical Studies in Oceanography, 116, pp.332-341.

Reddy, K.J., Haskell, J.B., Sherman, D.M. and Sherman, L.A., 1993. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. Journal of Bacteriology, 175(5), pp.1284-1292.

Redfield, A.C., 1963. The influence of organisms on the composition of sea-water. The sea, pp.26-77.

Rees, A.P., Gilbert, J.A. and Kelly-Gerreyn, B.A., 2009. Nitrogen fixation in the western English Channel (NE Atlantic ocean). Marine Ecology Progress Series, 374, pp.7-12.

Riemann, L., Farnelid, H. and Steward, G.F., 2010. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. Aquatic Microbial Ecology, 61(3), pp.235-247.

Sargent, E.C., Hitchcock, A., Johansson, S.A., Langlois, R., Moore, C.M., LaRoche, J., Poulton, A.J. and Bibby, T.S., 2016. Evidence for polyploidy in the globally important diazotroph *Trichodesmium*. FEMS Microbiology Letters, 363(21), p.fnw244.

Schlesier, J., Rohde, M., Gerhardt, S. and Einsle, O., 2015. A Conformational Switch Triggers Nitrogenase Protection from Oxygen Damage by Shethna Protein II (FeSII). Journal of the American Chemical Society, 138(1), pp.239-247.

Schlitzer, R., 2015. Ocean Data View, odv. awi. de.

Schunck, H., Lavik, G., Desai, D.K., Großkopf, T., Kalvelage, T., Löscher, C.R., Paulmier, A., Contreras, S., Siegel, H., Holtappels, M. and Rosenstiel, P., 2013. Giant hydrogen sulfide plume in the oxygen minimum zone off Peru supports chemolithoautotrophy. PLoS One, 8(8), p.e68661.

Severin, I., Bentzon-Tilia, M., Moisander, P.H. and Riemann, L., 2015. Nitrogenase expression in estuarine bacterioplankton influenced by organic carbon and availability of oxygen. FEMS microbiology letters, 362(14), p.fnv105.

Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S. and Furuya, K., 2014. Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean. Global Biogeochemical Cycles, 28(10), pp.1096-1110.

Sieburth, J.M., 1987. Contrary habitats for redox-specific processes: methanogenesis in oxic waters and oxidation in anoxic waters. Microbes in the Sea, pp.11-38.

Singh, B.K. and Walker, A., 2006. Microbial degradation of organophosphorus compounds. FEMS microbiology reviews, 30(3), pp.428-471.

Smith, D.C., Simon, M., Alldredge, A.L. and Azam, F., 1992. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. Nature, 359(6391), pp.139-142.

Sobarzo, M., Bravo, L., Donoso, D., Garcés-Vargas, J. and Schneider, W., 2007. Coastal upwelling and seasonal cycles that influence the water column over the continental shelf off central Chile. Progress in Oceanography, 75(3), pp.363-382.

Sohm, J.A., Hilton, J.A., Noble, A.E., Zehr, J.P., Saito, M.A. and Webb, E.A., 2011a. Nitrogen fixation in the South Atlantic Gyre and the Benguela upwelling system. Geophysical Research Letters, 38(16).

Sohm, J.A., Subramaniam, A., Gunderson, T.E., Carpenter, E.J. and Capone, D.G., 2011b. Nitrogen fixation by *Trichodesmium* spp. and unicellular diazotrophs in the North Pacific Subtropical Gyre. Journal of Geophysical Research: Biogeosciences, 116(G3).

Stal, L.J., 2009. Is the distribution of nitrogen-fixing cyanobacteria in the oceans related to temperature?. Environmental microbiology, 11(7), pp.1632-1645.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 22(21), pp.2688-2690.

Stramma, L., Johnson, G.C., Sprintall, J. and Mohrholz, V., 2008. Expanding oxygen-minimum zones in the tropical oceans. science, 320(5876), pp.655-658.

Stramma, L., Prince, E.D., Schmidtko, S., Luo, J., Hoolihan, J.P., Visbeck, M., Wallace, D.W., Brandt, P. and Körtzinger, A., 2012. Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. Nature Climate Change, 2(1), pp.33-37.

Subramaniam, A., Mahaffey, C., Johns, W. and Mahowald, N., 2013. Equatorial upwelling enhances nitrogen fixation in the Atlantic Ocean. Geophysical Research Letters, 40(9), pp.1766-1771.

Summons, R.E., Jahnke, L.L., Hope, J.M. and Logan, G.A., 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature, 400(6744), pp.554-557.

Suzuki, M.T. and Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and environmental microbiology, 62(2), pp.625-630.

Swannell, R.P., Lee, K. and McDonagh, M., 1996. Field evaluations of marine oil spill bioremediation. Microbiological Reviews, 60(2), pp.342-365.

Tamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution, 24(8), pp.1596-1599.

Team, R.C., 2013. R: A language and environment for statistical computing.

Thamdrup, B., Dalsgaard, T. and Revsbech, N.P., 2012. Widespread functional anoxia in the oxygen minimum zone of the Eastern South Pacific. Deep Sea Research Part I: Oceanographic Research Papers, 65, pp.36-45.

Thompson, A., Carter, B.J., Turk-Kubo, K., Malfatti, F., Azam, F. and Zehr, J.P., 2014. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. Environmental microbiology, 16(10), pp.3238-3249.

Thorneley, R.N. and Ashby, G.A., 1989. Oxidation of nitrogenase iron protein by dioxygen without inactivation could contribute to high respiration rates of Azotobacter species and facilitate nitrogen fixation in other aerobic environments. Biochemical Journal, 261(1), pp.181-187.

Tu, Q., Deng, Y., Yan, Q., Shen, L., Lin, L., He, Z., Wu, L., Van Nostrand, J.D., Buzzard, V., Michaletz, S.T. and Enquist, B.J., 2016. Biogeographic patterns of soil diazotrophic communities across six forests in the North America. Molecular ecology.

Turk, K.A., Rees, A.P., Zehr, J.P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E.M.S. and Gilbert, J., 2011. Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. The ISME journal, 5(7), pp.1201-1212.

Turk-Kubo, K.A., Karamchandani, M., Capone, D.G. and Zehr, J.P., 2014. The paradox of marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. Environmental microbiology, 16(10), pp.3095-3114.

Turk-Kubo, K.A., Frank, I.E., Hogan, M.E., Desnues, A., Bonnet, S. and Zehr, J.P., 2015. Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon). Biogeosciences, 12(24), pp.7435-7452.

Udiković-Kolić, N., Scott, C. and Martin-Laurent, F., 2012. Evolution of atrazine-degrading capabilities in the environment. Applied microbiology and biotechnology, 96(5), pp.1175-1189.

van der Maarel, M.J., Sprenger, W., Haanstra, R. and Forney, L.J., 1999. Detection of methanogenic archaea in seawater particles and the digestive tract of a marine fish species. FEMS microbiology letters, 173(1), pp.189-194.

Vitousek, P.M. and Howarth, R.W., 1991. Nitrogen limitation on land and in the sea: how can it occur?. Biogeochemistry, 13(2), pp.87-115.

Vitousek, P.M., Menge, D.N., Reed, S.C. and Cleveland, C.C., 2013. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1621), p.20130119.

Voss, M., Bombar, D., Loick, N. and Dippner, J.W., 2006. Riverine influence on nitrogen fixation in the upwelling region off Vietnam, South China Sea. Geophysical Research Letters, 33(7).

Voss, M., Croot, P., Lochte, K., Mills, M. and Peeken, I., 2004. Patterns of nitrogen fixation along 10 N in the tropical Atlantic. Geophysical Research Letters, 31(23).

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S. and Schwartz, M., 2013. gplots: Various R programming tools for plotting data. R package version 2.12. 1.

Wasmund, N., Struck, U., Hansen, A., Flohr, A., Nausch, G., Grüttmüller, A. and Voss, M., 2015. Missing nitrogen fixation in the Benguela region. Deep Sea Research Part I: Oceanographic Research Papers, 106, pp.30-41.

Wintzingerode, F.V., Göbel, U.B. and Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS microbiology reviews, 21(3), pp.213-229.

Witty, J.F. and Minchin, F.R., 1998. Hydrogen measurements provide direct evidence for a variable physical barrier to gas diffusion in legume nodules. Journal of Experimental Botany, 49(323), pp.1015-1020.

Xiao, P., Jiang, Y., Liu, Y., Tan, W., Li, W. and Li, R., 2015. Re-evaluation of the diversity and distribution of diazotrophs in the South China Sea by pyrosequencing the nifH gene. Marine and Freshwater Research, 66(8), pp.681-691.

Yogev, T., Rahav, E., Bar-Zeev, E., Man-Aharonovich, D., Stambler, N., Kress, N., Béjà, O., Mulholland, M.R., Herut, B. and Berman-Frank, I., 2011. Is N_2 fixation significant in the Levantine Basin, east Mediterranean Sea?. Environmental microbiology, 13(4), pp.854-871.

Zehr, J.P., 2011. Nitrogen fixation by marine cyanobacteria. Trends in microbiology, 19(4), pp.162-173.

Zehr, J.P., Crumbliss, L.L., Church, M.J., Omoregie, E.O. and Jenkins, B.D., 2003b. Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of functional genes. Biotechniques, 35(5), pp.996-1013.

Zehr, J.P., Jenkins, B.D., Short, S.M. and Steward, G.F., 2003a. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental microbiology, 5(7), pp.539-554.

Zehr, J.P., Mellon, M.T. and Zani, S., 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes. Applied and environmental microbiology, 64(9), pp.3444-3450.

Zehr, J.P., Montoya, J.P., Jenkins, B.D., Hewson, I., Mondragon, E., Short, C.M., Church, M.J., Hansen, A. and Karl, D.M., 2007. Experiments linking nitrogenase gene expression to nitrogen fixation in the North Pacific subtropical gyre. Limnology and Oceanography, 52(1), pp.169-183.

Zehr, J.P., Bench, S.R., Carter, B.J., Hewson, I., Niazi, F., Shi, T., Tripp, H.J. and Affourtit, J.P., 2008. Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. Science, 322(5904), pp.1110-1112.

Zhao, G. and Winkler, M.E., 1996. A novel alpha-ketoglutarate reductase activity of the serA-encoded 3-phosphoglycerate dehydrogenase of Escherichia coli K-12 and its possible implications for human 2-hydroxyglutaric aciduria. Journal of bacteriology, 178(1), pp.232-239.

CHAPTER 5: ISOLATION AND GENOME SEQUENCING OF A NOVEL MARINE HETEROTROPHIC DIAZOTROPH SHEDS LIGHT ON ITS LIFESTYLE

Jenni-Marie Ratten, André Comeau, Jennifer Tolman, Morgan Langille, Julie LaRoche

Contribution of authors:

Jenni-Marie Ratten:	Isolation and cultivation of organism, data analysis, drafting of the manuscript
André Comeau:	Sequencing and assembly of the genome
Jennifer Tolman:	Technical support during cell sorting flow cytometry and technical support with qPCR assay
Morgan Langille:	Sequencing and assembly of the genome
Julie LaRoche:	Planning and discussion of the manuscript

5.0. Abstract

A novel heterotrophic diazotroph from the order Oceanospirilalles (gamma-Proteobacteria) was isolated by enrichment culture and single-cell sorting from the surface waters of Bedford Basin, an oceanic inlet (Halifax, Canada). Based on the 16S rRNA gene, the isolate's closest relative is *Thalassolituus oleivorans* R6-14 (97.2% sequence similarity); however, *T. oleivorans* is not a diazotroph. The closest diazotrophic relative based on full length *nifH* sequence is *Marinobacterium litorale* (84.7% sequence similarity). Quantitative PCR (qPCR) showed that the new isolate is widely distributed throughout the temperate North Atlantic along the GEOVIDE cruise transect (spanning from Portugal to Labrador via Greenland) and throughout the Scotian Shelf. We found a maximum of 2.5 x 10^5 *nifH* gene copies L^{-1} in the Bedford Basin on February 26, 2014 at 10 m water depth. A database search showed that the isolate was present in the deep Pacific and in the Chilean oxygen minimum zone (OMZ). Genome sequencing using both Illumina whole-genome shot-gun and Nanopore MinION technology resulted in a genome of 4.44 Mbp with 60 RNAs and 4003 coding regions comprising 463 subsystems. A complete *nif* operon and several other gene clusters characteristic of diazotrophs were annotated indicating that the isolate has the potential to perform N_2 fixation. Revers-transcriptase PCR of *nifH* transcripts from cells grown in nitrogen-depleted media suggest active expression of the nitrogenase enzyme. The novel strain grew best when supplemented with nitrate and a carbon source such as acetate or Tween20. Although laboratory conditions were aerobic, the isolate's genome suggests that it can carry out anaerobic respiration, lactatic acid

fermentation and use a range of mechanisms to deal with oxidative stress to prevent oxidative damage to the nitrogenase. The isolate also possesses genes for the uptake of several trace metals, secretion transporters to establish optimal ion balance, and aromatic carbon degradation to deal with potentially toxic metals and compounds. With the emergence of evidence for the importance of non-cyanobacterial diazotrophs and a lack of cultured representatives, this isolate provides some insight into the metabolism of the still vastly unknown assemblage of marine non-cyanobacterial diazotrophs.

5.1. Introduction

Marine N_2 fixation has long been attributed to cyanobacterial diazotrophs (Zehr 2011). However, studies using the *nifH* gene as a functional marker for diazotrophs have revealed that the bulk of the diazotrophic diversity resides within the non-cyanobacteria dominated cluster I (Chapter 3 and 4; Bombar et al. 2016; Farnelid et al. 2011; Zehr et al. 2003). Although many of these potential heterotrophic diazotrophs have only been identified as *nifH* phylotypes, they can be dominant representatives in the temperate open ocean and other specific niches such as the oxygen minimum zones (OMZ; Langlois et al. 2015; Loescher et al. 2014; Fernandez et al. 2011; Rieman et al. 2010; Bird et al. 2005). Proteobacteria in general, exhibit very diverse metabolic potential, and the marine component of this phylum is likely to show similar metabolic diversity. Physiological studies on three recently isolated diazotrophs from a brackish habitat reinforced this view, showing that optimal growth conditions varied widely

for the three unrelated strains within the proteobacterial cluster (Bentzon-Tilia et al. 2015). To determine the functional role of non-cyanobacterial marine diazotrophs, more cultured laboratory strains are needed to study their metabolism in controlled conditions. In this study, we report the isolation, cultivation and genome sequence of a novel marine heterotrophic diazotroph first isolated from the Bedford Basin, a temperate coastal Atlantic Ocean inlet in Nova Scotia, Canada. The isolate is widely distributed throughout the North Atlantic Ocean from the coast of Portugal to the Canadian east coast. The wide geographical distribution of this isolate was confirmed both by high-throughput sequencing of *nifH* amplicons and by a phylotype-specific qPCR assay, that detected the specific *nifH* gene at concentrations reaching up to 2.5×10^5 *nifH* copies L^{-1}, which is three times higher than the maximum *nifH* concentrations recorded for another widespread gamma-proteobacterial diazotroph (GammaA: 8.0×10^4 *nifH* copies L^{-1}; Langlois et al. 2015). Sequences with 99% similarity to the isolate's *nifH* gene have been previously reported from the deep North Pacific Ocean and the Chilean OMZ (Fernandez et al. 2011; Mehta et al. 2005).

The genome sequence has allowed us to gain some insight into its potential lifestyle, including diazotrophy, iron uptake, strategies to deal with O_2 stress and its ability to catabolise various complex carbon sources.

5.2. Methods

5.2.1. Sample collection and sea water enrichments

Water samples were collected at the Compass Buoy station in the Bedford Basin (44^0 41' 30'' N, 63^0 38' 30''W) at depths of 1, 5, 10 and 60 m on January 29, 2014. Sampling was performed as part of the Bedford Basin Monitoring Program of the Bedford Institute of Oceanography (BIO). 30 mL aliquots of the samples were enriched with nutrients according to Table 5.1.

Table 5.1: Enrichment treatments for diazotrophic isolation.

	NH_4NO_3 (2 μM)[1]	Phosphate (400nM)	Iron(II) (4nM)	Glucose (2 μM)	Thiosulfate (1 mM)
Control	-	-	-	-	-
1.	-	+	+	-	-
2.	-	+	+	+	-
3.	-	+	+	+	+
4.	-	-	-	-	+
5.	+	+	+	-	-

1) Final concentration in sample

Enrichments from 1, 5 and 10 m depth were incubated at both 12 and 5°C with 12-hour light/dark cycles. The 60 m enrichments were incubated at 4°C in the dark to resemble conditions at 60 m in the Bedford Basin. The incubated flasks were re-enriched monthly and 10 µL aliquots were simultaneously monitored for the presence of *nifH* (all PCR steps are described in the PCR amplification section). One enrichment tested positive for *nifH* (phosphate and iron enrichment, 10 m depth) and was further processed. Fluorescent Activated Cell Sorting

(FACS) was used to subdivide the populations of the enrichment with the goal of enriching a diazotroph fraction within the sample; *nifH* PCR was performed on 200 cells sorted from each population into 10 µL PCR water. Single cells from sub-populations that tested positive for *nifH* were sorted on f/2 artificial sea water agar plates (1.2%; Guillard and Ryther 1962). Plates were incubated according to the conditions of the original enrichment culture (5°C, 12-hour dark/light cycle). Colonies from single cell sorts formed within 90 days, and were screened for *nifH* gene presence by colony PCR. For DNA extraction, positive colonies were inoculated into liquid YBCII medium amended with Sodium Acetate (15 mM; Chen at al. 1996). Colonies were regrown on artificial sea water agar plates (1.2%) for Transmission Electron Microscopy preparation.

5.2.2. DNA/RNA extraction and reverse transcription

DNA and RNA were extracted using the AllPrep Mini Kit (Qiagen) according to the manufacturer's instructions, with the following modifications: 100 mL of culture from one of the *nifH* positive clones was grown to visible density in YBCII medium supplemented with 15 mM sodium acetate and cells were pelleted at 14,000 rpm, resuspended in 50 µL lysozyme solution (20 mg/mL in TE buffer) and incubated for 5 min. Next, 45 µL Proteinase K and 600 µL RLT buffer with 10 µL β-mercaptoethanol were added. After incubation at 52°C for 15 min, the mixture was passed through a Qiashredder column (Qiagen). Further extraction was performed according to the manufacturer's protocol. DNA was eluted in 50 µL TE buffer and RNA in 50 µL RNase-free water. Reverse transcriptions was

done using random hexamers and SuperScript® III Reverse Transcriptase (Invitrogen) including no-template controls according to the manufacturer's instructions.

5.2.3. PCR amplification

Cells from 10 µL aliquots of enrichment cultures, population sorts or plate-grown colonies were lysed by 3 repeated freezing (-20°C, 3 hours) and heating (100°C, 5 min) cycles. The presence of *nifH* genes was also tested in RT-PCR samples with RNA-template controls. PCR was performed according to the nested protocol of Zehr et al. (2001). The first amplification contained 5 µL 10x buffer (QIAGEN), 4 µL 10 µM dNTPs (Invitrogen), 8 µL 25 mM $MgCl_2$ (QIAGEN), 4 µL 10 µM of each nifH3 and nifH4 primers, 0.6 µL BSA (20mg/mL; NEB) 10 µL template, 0.25 µL QIAGEN HotStar Taq polymerase (1.25 U) and 14.15 µL PCR grade water for a 50 µL final volume. The PCR cycling protocol was: 95°C for 15 min followed by 35 cycles of 95°C (1 min), 45°C (1 min) and 72°C (1 min) with a final 10 min extension at 72°C. The second amplification was a 10 µL reactions with 1 µL 10x buffer, 0.8 µL 10 µM dNTPs, 1.2 µL $MgCl_2$, 0.8 µL of each nifH1 and nifH2 primers, 0.06 µL BSA, 1 µL PCR template from the first amplification, 0.05 µL Qiagen HotStar Taq polymerase (0.25 U) and 4.29 µL PCR water. Cycling for the second PCR followed: 95°C for 15 min, 28 cycles at 95°C (1 min), 54°C (1 min) and 72°C (1 min), and the final extension at 72°C for 10 min. No-template controls were included in all PCR assays.

5.2.4. Quantitative PCR

Primers (for: 5'-AGCCCGGTGTTGGTTGTG-3', rev: 5'-AAGCACCTTCTTCTTCGAGGAA-3'; IDT) and probe (6FAM-TCGCGGTGTCATCACAGCGATCA; Applied Biosystems) for TaqMan *nifH* qPCR were designed using Primer Express (version 3.0; Applied Biosystems). BLAST was used against the non-redundant (nr) database to ensure that neither primers nor probe targeted any other sequences (Altschul et al. 1990). An oligomer of the probe sequence was used as a standard. Duplicate standards and individual sample qPCR reactions were run on a StepOnePlus (Applied Biosystems). The 18 µL reaction contained 9 µL TaqMan Universal PCR Master Mix (Applied Biosystems), 3 µM forward primer, 1 µM reverse primer, 100 nM probe, and 1 µL DNA template. PCR grade water was used as a no-template control. Cycling conditions were: 50°C (2 min), 95°C (10 min) and 45 cycles of 95°C (15 s) and 60°C (1 min). *NifH* gene copies L^{-1} of the isolated organism were calculated for 549 qPCR samples collected at the Compass Buoy station in the Bedford Basin in 2014 (181 samples), during two Scotian Shelf research cruises (AZMP HUD2014004 and HUD2014030; 164 samples) and the GEOVIDE research cruise spanning from Portugal to Labrador via southern Greenland (204 samples).

5.2.5. Microscopy

A colony of the isolate grown on artificial sea water was smeared on a glass slide, heat-fixed, Gram-stained and observed using a Zeiss microscope Imager.M2 with Apotome.2.

For transmission electron microscopy, a colony of isolated diazotroph cells growing on artificial seawater was directly fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer. Further processing was conducted by the electron microscope facility of Dalhousie University (Halifax, Canada). The sample was sequentially rinsed 3 times with 0.1 M sodium cacodylate buffer, fixed for 2 hours with 1% osmium tetroxide, rinsed with distilled water, and stained with 0.25% uranyl acetate at 4°C overnight. The sample was then dehydrated with a graduated series of acetone/H_2O mixtures: 50% acetone (10 min), 70% acetone (10 minutes twice), 95% acetone (10 min twice), 100% acetone (10 minutes twice), and anhydrous acetone (10 min). Then, the sample was gradually infiltrated with Epon-Araldite resin: 3 parts anhydrous acetone : 1 part resin (3 hours), 1 part anhydrous acetone : 3 parts resin (overnight) and 100% Epon-Araldite resin (3 hours twice). After embedding in 100% Epon-Araldite resin, the sample was incubated at 60°C for 48 hours to cure. Ultrathin sections were cut using a Reichert-Jung Ultracut E Ultramicrotome with a diamond knife and placed on 300 mesh copper grids. Staining was performed as follows: 2% aqueous uranyl acetate (10 min), two rinses with distilled water, lead citrate (4 min), rinse with distilled water, and air-dry. Images were taken with an FEI Tecnai-12 and a

Gatan 832 CCD camera at the Scientific Imaging Suite of the Biology Department at Dalhousie University (Halifax, Canada).

5.2.6. Genome Sequencing

Genomic DNA was sequenced both on an Illumina MiSeq and a Nanopore MinION device through the Integrated Microbiome Resource (IMR) of the Centre for Comparative and Evolutionary Biology (CGEB) at Dalhousie University (Halifax, Canada). For Illumina sequencing, the library was constructed using the Illumina Nextera XT kit directly from 1 ng of DNA, dual-indexed, then run on a MiSeq using v3 600 cycle chemistry (300+300 bp). For Nanopore sequencing, 1 µg of DNA was sheared using a Covaris g-TUBE (5000 x g for 1 min, then 1 min recovery spin), the library was constructed using the Nanopore SQK-MAP006 kit (without the optional repair step), then run on one R7 flowcell on a Mk1 MinION device (48 h protocol).

Over 500 Mb of Illumina MiSeq data, 113 Mb of MinION "1D pass" and 26 Mb of "2D pass" data were generated and then used to evaluate the performance of standard (Illumina or MinION only) and hybrid (Illumina+MinION) assemblers, including: A5 (Tritt et al. 2012), Canu (Koren et al. 2017), Celera Assembler (Myers et al. 2000), LINKS (Warren et al. 2015), newbler (Roche AG), SPAdes (Nurk et al. 2013), and SSPACE-LongRead (Boetzer and Pirovano 2014). Two MinION read correction programs (Nanocorr and NaS; Goodwin et al. 2015; Mandoui et al. 2015) were also evaluated. Traditional, Illumina-only assemblies generated ~50 contigs/scaffolds (total assembly size of 4.44 Mb), whereas hybrid

assembly with uncorrected MinION data decreased this number to an average of ~20. Best results were obtained in hybrid assemblies with corrected MinION reads, generating <10 scaffolds converging upon a final assembly length of 4.44 Mb with median coverage of 38 and N50 of ~2.4 Mb. The final best combination was MiSeq data + MinION raw 2D data in the SPAdes assembler, followed by scaffolding in SSPACE-LongRead using the output SPAdes contigs + 1D raw + 2D-NaS-corrected MinION data, which created 4 scaffolds. Some final gap closure was undertaken with FGAP (T100, R10k, I10k parameters; Piro et al. 2014) using the same data as for SSPACE. A few spurious contigs and/or breaks in assemblies were caused by rRNA operons and mobile elements.

5.2.7. Phylogenetic analysis

To investigate the phylogenetic affinities of the isolate, the 16S rRNA gene sequences of the 12 closest reference genomes, and 132 diazotrophic reference genomes were extracted from NCBI. 16S rRNA gene alignments of the closest reference genomes and the diazotrophic reference genomes were constructed using MAFFT v. 7 (Yamada et al. 2016; Katoh et al. 2002) and by romoving ambiguous sequence alignment regions using Gblocks (Castresana 2000). Maximum likelihood analysis was performed using RAxML with the GTR-GAMMA model (Stamakatis 2014). The *nifH* genes of the isolate and the diazotrophic reference genomes were also extracted, aligned based on the protein sequence using MAFFT v. 7, returned to nucleotide sequences with PAL2NAL, removing ambiguos sequence alignment regions with Gblocks and a maximum likelihood

tree was inferred using RAxML with the GTR-GAMMA model and bootstrap values were calculated from 100 replicates (Yamada et al. 2016; Stamakatis 2014; Suyama et al. 2006; Katoh et al. 2002; Castresana 2000). All trees were displayed using iTOL (interactive Tree of Life; Letunic and Borg 2016).

5.2.8. Statistical Analysis of correlation of abundances with environmental parameters

Statistical analyses were performed using PRIMER-E version 6.1.12 (Clarke and Gorley, 2006). Out of the 549 analysed samples, hydrographic measurements and environmental parameters were available for 436 samples, which included depth, temperature, salinity, O_2 concentration, nitrate, nitrite, ammonium, silicate and chlorophyll concentrations (personal communication with the Bedford Institute of Oceanography and Richard Davis, CERC.OCEAN Dalhousie University and the GEOVIDE LKEF-CYBER database; Supplemental Table 10). Both the environmental matrix and the *nifH* abundance matrix obtained from qPCR were log-transformed. Bray-Curtis similarities were generated for diazotrophic abundances. A BEST (Bio-Env + Stepwise) test identified the environmental factors that best explain abundance distribution. A Principle Component Analysis (PCA) was performed with the environmental matrix to find environmental parameters that may influence the isolate's distribution. The first three components of the PCA captured 77.7% (PC1 41.0%, PC2 23.8%, PC3 12.9%) of the variance in the environmental parameters of the Bedford Basin, AZMP HUD2014004 and HUD2014030 and GEOVIDE cruises.

5.3. Results

5.3.1. Physiological traits

This novel heterotrophic diazotroph was isolated from 10 m depth in the Bedford Basin on January 29, 2014. Light microscopy examination of a Gram stain revealed that this species is Gram-negative. Transmission electron microscopy of a fixed colony grown on sea water agar showed that the cells were rod-shaped, approximately 2.5 µm long and 0.5 µm in diameter (Figure 5.1). The nucleoid was located in the centre of the cytoplasm. Most cells also contained electron-translucent inclusions at each cell pole (Figure 5.1).

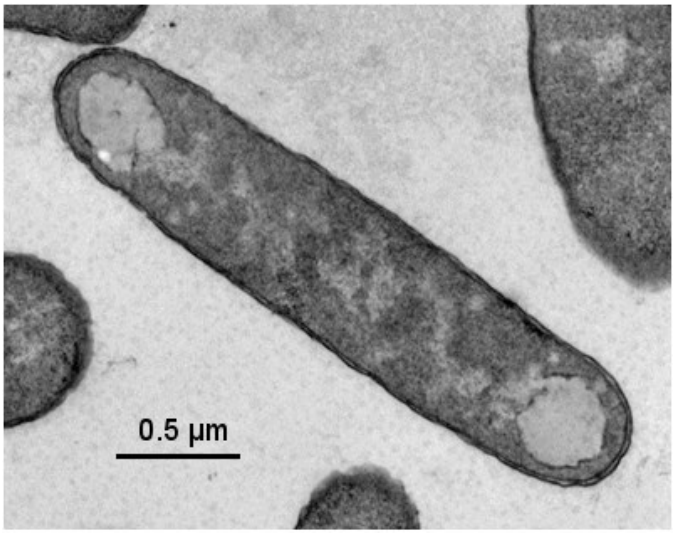


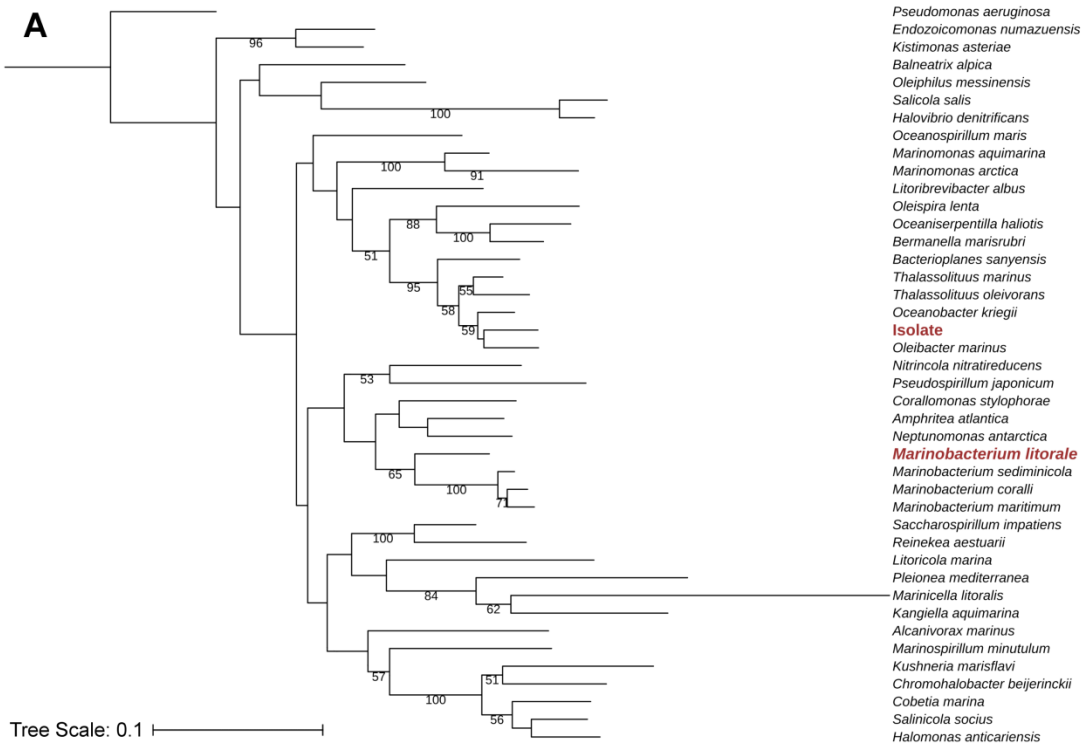
Figure 5.1: Transmission Electron Microscopy image of the isolated diazotroph. Transmission electron micrograph of the isolate, using an FEI Tecnai-12 and a Gatan 832 CCD camera. The scale bar represents 0.5 µm. Inclusions at cell poles can be seen.

The isolate tended to flocculate when grown in liquid culture. In growth experiments, optimal conditions were achieved in artificial sea water supplemented with the supplementation of acetate and nitrate (Supplemental Figure 12). The capability for N_2 fixation was shown indirectly by demonstrating the presence of *nifH* transcription in cells grown in nitrogen-depleted medium (Supplemental Figure 13).

5.3.2. Genomic analysis

Illumina MiSeq shotgun sequencing followed by assembly resulted in 48 contigs with a total length of 4.36 Mb. Mapping these contigs onto long reads produced by Nanopore's MinION reduced the contig number to 4 with a length of 4.44 Mb. The genome was annotated using SEED, which found 4003 coding regions of 463 subsystems and 60 RNAs (Overbeek et al. 2006). The GC content of the genome was 53.4% (Table 5.2).

The *nifH* and 16S rRNA genes sequences were used to further explore the taxonomic assignment of this new isolate. Following alignment and phylogenetic analysis with 16S rRNA and *nifH* gene reference sequences, the isolate was placed in a clade among the gamma-proteobacteria (Figure 5.2). The closest relatives are *Marinobacterium litorale* (based on its *nifH* sequence, 84.7% identity) and *Thalassolituus oleivorans* R6-14 (based on the 16S rRNA gene, 97.2% identity), suggesting that it is a member of the order of Oceanospirilalles.


A
96
100
100
91
88
100
51
95
55
58
59
53
65
100
71
100
84
62
57
51
100
56
Tree Scale: 0.1
Pseudomonas aeruginosa
Endozoicomonas numazuensis
Kistimonas asteriae
Balneatrix alpica
Oleiphilus messinensis
Salicola salis
Halovibrio denitrificans
Oceanospirillum maris
Marinomonas aquimarina
Marinomonas arctica
Litoribrevibacter albus
Oleispira lenta
Oceaniserpentilla haliotis
Bermanella marisrubri
Bacterioplanes sanyensis
Thalassolituus marinus
Thalassolituus oleivorans
Oceanobacter kriegii
Isolate
Oleibacter marinus
Nitrincola nitratireducens
Pseudospirillum japonicum
Corallomonas stylophorae
Amphritea atlantica
Neptunomonas antarctica
Marinobacterium litorale
Marinobacterium sediminicola
Marinobacterium coralli
Marinobacterium maritimum
Saccharospirillum impatiens
Reinekea aestuarii
Litoricola marina
Pleionea mediterranea
Marinicella litoralis
Kangiella aquimarina
Alcanivorax marinus
Marinospirillum minutulum
Kushneria marisflavi
Chromohalobacter beijerinckii
Cobetia marina
Salinicola socius
Halomonas anticariensis

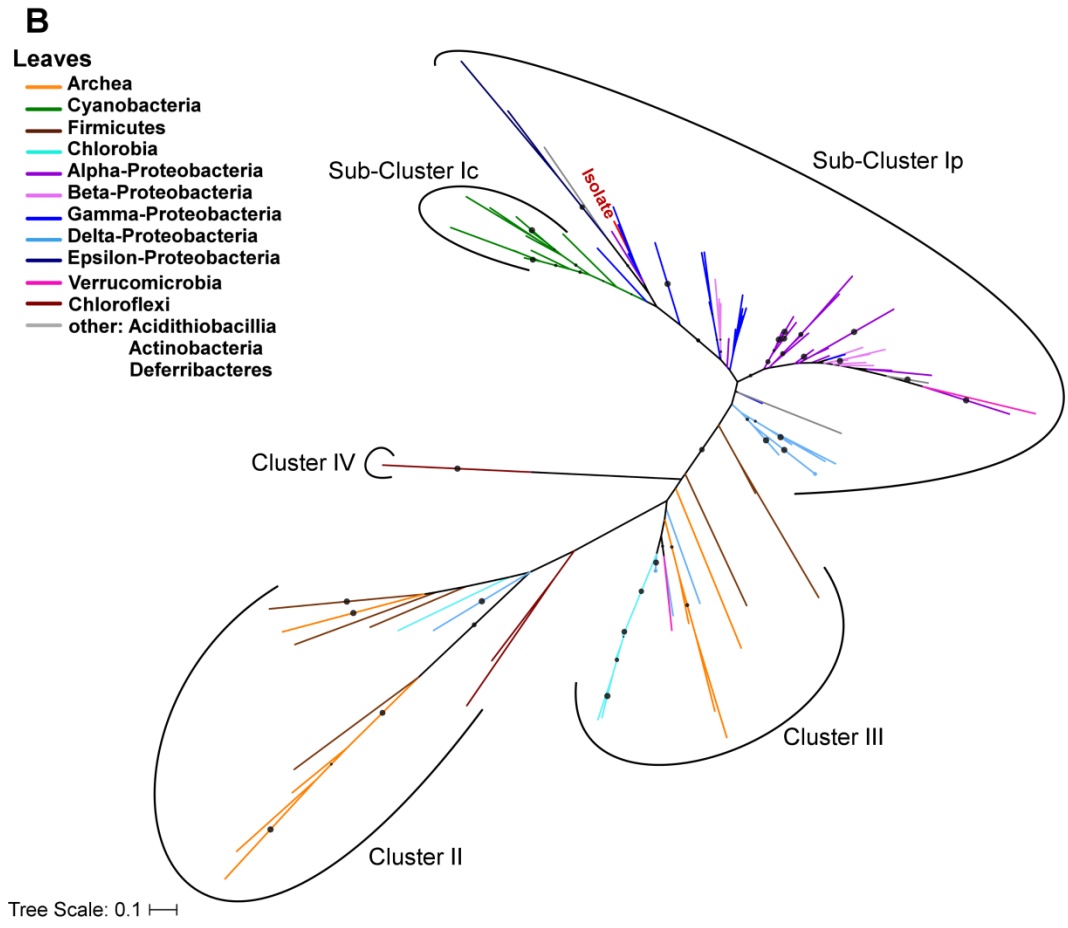

B
Leaves
Archea
Cyanobacteria
Firmicutes
Chlorobia
Alpha-Proteobacteria
Beta-Proteobacteria
Gamma-Proteobacteria
Delta-Proteobacteria
Epsilon-Proteobacteria
Verrucomicrobia
Chloroflexi
other: Acidithiobacillia
Actinobacteria
Deferribacteres
Sub-Cluster Ic
Isolate
Sub-Cluster Ip
Cluster IV
Cluster III
Cluster II
Tree Scale: 0.1

Figure 5.2: Phylogenetic affiliation of the isolated diazotroph.

The phylogenetic affiliation of the isolate was established through DNA alignment of 16S rRNA genes, protein alignment of *nifH* sequences in MAFFT and consequent maximum likelihood tree building with nucleotide sequences using the GTR-GAMMA model in RAxML (Yamada et al. 2016; Stamakatis 2014; Suyama et al. 2006; Katoh et al. 2002). Bootstrap values were calculated from 100 tree replicates and values >50% are shown as dots. The trees - A: closest 16S rRNA reference genomes, B: diazotrophic reference genomes (*nifH* genes) - were displayed with branch lengths showing the number of substitutions per site as indicated for each tree. Branch colour indicates taxonomy according to panel B. The *nifH* tree was segmented into clusters I – IV (Zehr et al. 2003).

Table 5.2: Selected features of the sequenced genome as annotated by SEED (Overbeek et al. 2005).

Genome Features	Amount
Genome size (Mb)	4.44
% GC content	53.4
No. of subsystems	464
No. of coding sequences	4003
No. of RNA genes	60
Mobile elements	12
Nitrogen Metabolism	**69**
N_2 fixation	26
Nitrate and nitrite ammonification	14
Ammonia assimilation	24
Nitrosative stress	2
Cyanate hydrolysis	3
Iron Metabolism	**26**
Iron acquisition and metabolism	13
Heme, hemin uptake and utilization systems in Gram Positives	2
Transport of Iron	11
Oxygen Metabolism	**154**
Anaerobic respiratory reductases	28
Oxidative stress	53
Dioxygenases	15
Fermentation	58
Aromatic Compound Metabolism	**127**
Salicylate ester degradation	5
Phenol hydroxylase	12
Quinate degradation	1
Biphenyl Degradation	16
Benzoate degradation	11
p-Hydroxybenzoate degradation	1
Catechol branch of beta-ketoadipate pathway	3
Salicylate and gentisate catabolism	16
4-Hydroxyphenylacetic acid catabolic pathway	16
N-heterocyclic aromatic compound degradation	2
Central meta-cleavage pathway of aromatic compound degradation	25
Aromatic Amin Catabolism	9
Gentisate degradation	10

The two sections of the *nif* operon spread over a region of ca. 0.21 Mb and were separated by ca. 0.14 Mbp (Figure 5.3A). One section of the operon contains genes coding for the transcriptional regulator *nifA*, three structural units of the nitrogenase enzyme (*nifHDK*), FeMo cofactor synthesis supporting proteins (*Avin2460*, *nifB*, *nifQ*), and *nif* genes of unknown function. The other section contains coding regions that support the synthesis of the nitrogenase enzyme and its co-factors (*frdN*, *nifE*, *nifN*, *nifS*, *nifU*, *nifV*, *nifX*), the nitrogenase stabilizing protein *nifW*, and other *nif* genes of unknown function. The region between the *nif* clusters contains genes for glycogen metabolism. The GC content of the *nif* operon region amounted to 53.9% (cluster containing *nifHDK*: 53.1%, cluster containing support *nif* genes: 54%, region between clusters: 54.5%), while the sections surrounding the operon have a higher GC content of 55.5%.

Mapping genes to their function in the complete network of nitrogen-related metabolic pathways revealed that this isolate can produce and assimilate ammonium via several pathways in addition to N_2 fixation (Figure 5.3B). Pathways forming the products of allantoate (purine metabolism) and urea (arginine metabolism) can be utilized by this organism to remineralize nitrogen. The detoxification of cyanate and the reduction of nitrate are also possible pathways to the acquisition of ammonium (Figure 5.3B).

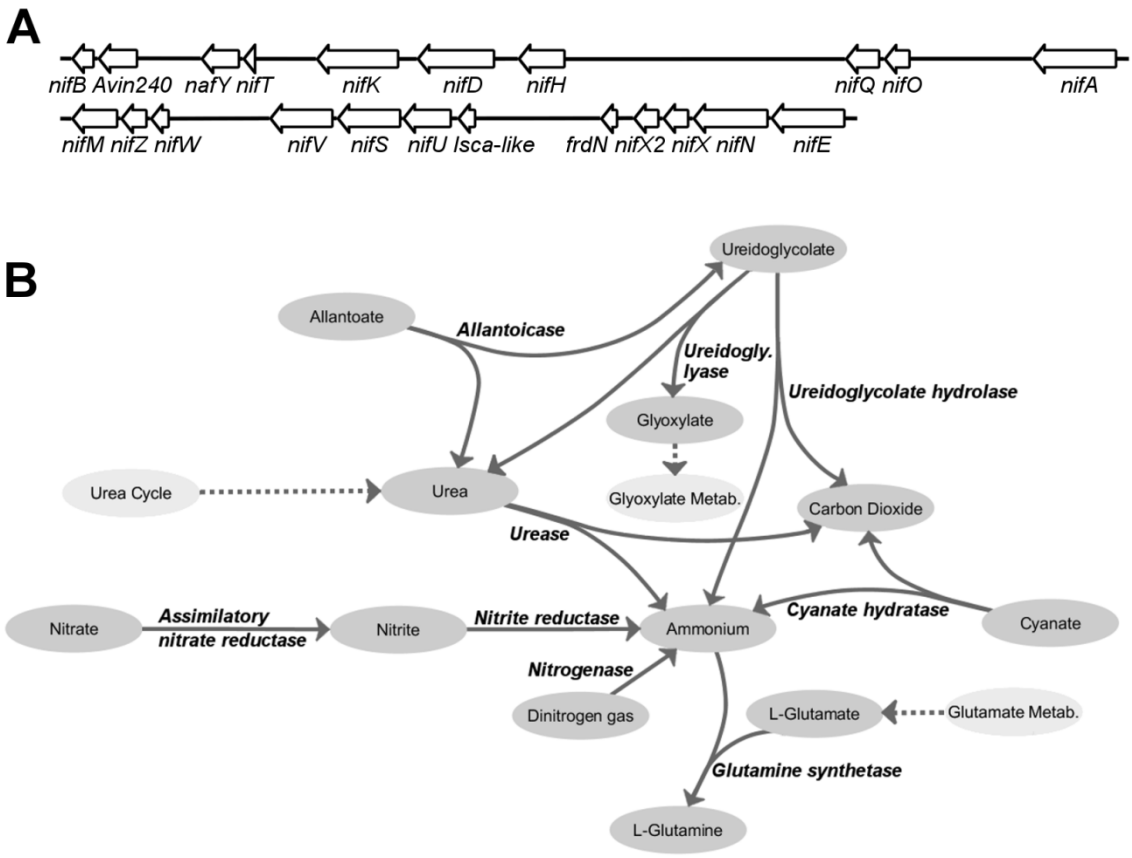


Figure 5.3: Nitrogen cycling associated pathways in the novel diazotroph isolate. The genome of the isolated diazotroph was annotated in the SEED database (Overbeek et al. 2005). Panel (A) shows the two *nif* operon regions (drawn to scale in Cytoscape 3.4.0 (Shannon et al. 2003)); the potential pathways for nitrogen metabolism as extracted from SEED are depicted in panel (B; created in Cytoscape 3.4.0 (Shannon et al. 2003)).

The genome provided physiological insight into this organism, beyond that already observed from culturing (Table 5.2 and Figure 5.4). Genes for a flagellum are present, suggesting that the organism is likely motile; chemotactic receptors and signalling pathways suggest that it is able to move along chemical gradients. Furthermore, genes for both aerobic respiration and anaerobic fermentation of acetyl-coA to butanol and pyruvate to lactate are present. There is a variety of

transporters for biopoly-, di- and monomers present to obtain macronutrients, as well as alternative metabolic pathways into the TCA cycle and glycolysis, pointing to the use several carbon substrates. Most of the subsystems for carbon degradation are dedicated to the breakdown of aromatic carbon compounds including phenols, catechols, salicylate esters, benzoates, xylenols and biphenols, as well as more complex nitrogen-containing aromatic compounds (Table 5.2). Pathways for carbon anabolism include: amino acids, proteins, carbohydrates, nucleotides, fatty acids, phospholipids, gram-negative cell wall synthesis and the synthesis of polyhydroxybutyrates (Akaraonye et al. 2010).

The pathways dealing with the reduction of oxidative stress and with iron metabolism are of specific interest, because each nitrogenase requires at least 20 iron atoms and its reactive site is highly susceptible to oxidation (Berman-Frank et al. 2007, 2001; Kustka et al. 2003; Orma-Johnson 1985). This organism contains a range of mechanisms involved in both the reduction of oxidative stress (peroxidases, glutaredoxins and the broader glutathione redox metabolism, superoxide dismutase, catalase, and nitric oxide dioxygenase) and the direct removal of O_2 (rubredoxins and dioxygenases). To address higher iron requirements, the isolate contains several iron uptake and regulation mechanisms: ferrous iron transport protein A and B (*feoAB*), ferrous iron sensing transcriptional regulator (*feoC*), ferric iron ABC transporter (*fbpABC*), iron-uptake factor B and C (*piuBC*), ferrichrome-iron receptor (*OMR1*), ferric uptake regulation protein (*fur*), iron-dependent repressor (*ideR*/*dtxR*), iron-responsive repressor (*rirA*), iron-responsive regulator (*irr*) and iron regulated protein A

precursor (*irpA*). It also contains heme uptake and processing pathways: hemin transport protein (*chu*) and heme oxygenases (*hem*, *hmu*, *idi*, *hyp*). Free iron and heme can increase the stress caused by reactive O_2 via the Fenton reaction, which is counteracted by the bacteriophytochrome heme oxygenase, the ferroxidase and iron uptake metabolism (Cabiscol et al. 2010).

In addition to highly-developed iron uptake machinery, there are several trace metal importers and exporters, including general ABC and Ton/Tol transporters as well as specific ion transporters present. This allows the export or balance of the potentially toxic metals As, Cd, Co, Cr, Cu, Hg, Mg, Na, Ni, Se and Zn.

As part of fulfilling their nutrient requirements, bacteria compete with other members of the microbial community for resources. Mechanisms deployed by bacteria include phage defense, out-competition via antibacterial molecules, and the acquisition of new genes through the uptake of foreign DNA. This isolate contains the CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated system) system in its genome, which acts as the bacterial immune system against phages (Doudna and Charpentier 2014). As part of its virulence, this organism possesses genes to secrete colicin, a bacteriocin that is taken up by other bacteria and can be cytotoxic through the depolarisation of the cytoplasmic membrane (Parker et al.1989). The BarA-UvrY system regulates stress response in bacteria and initiates various downstream signalling events that lead to more adaptable phenotype including genes for the uptake and integration of foreign DNA (DNA receptors, integrases and recombinases and Type IV Pili; Sahu et al. 2003). The genome contains antibiotic

resistances to tetracycline, beta-lactams and fluoroquinolones, which were likely taken up through one of those mechanisms.

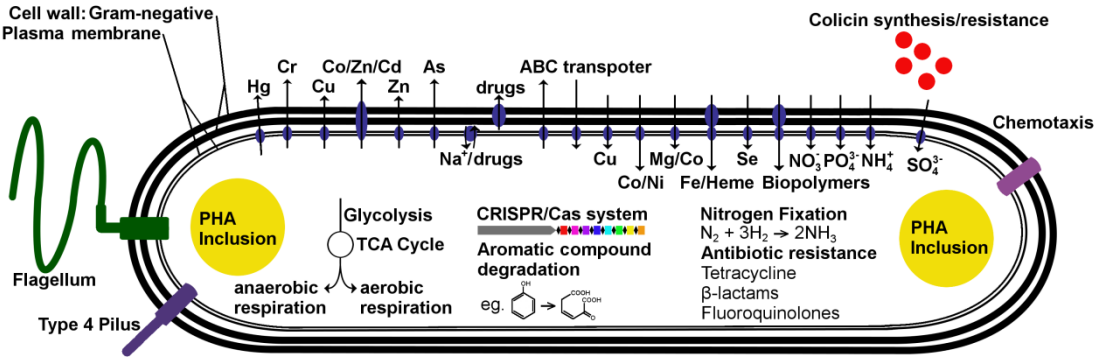


Figure 5.4: Metabolic reconstruction of the diazotrophic isolate.

The genome of the isolated diazotroph was annotated by SEED (Overbeek et al. 2005). Aspects of the assigned metabolic potential are highlighted here. Along with genes for glycolysis, the TCA cycle, anaerobic respiration and N_2 fixation, this organism has the potential to degrade various aromatic compounds, detoxify some antibiotics, transport a suite of toxic metals out of the cell, import many macro- and micronutrients, produce the bactericide colicin, exchange genetic information over a Type 4 pilus, and store carbon in the form of polyhydroxyalkanoates (PHA); it is likely motile, can follow a chemical gradients, and contains the bacterial immune system (CRISPR/Cas system).

5.3.3. Distribution and environmental conditions

Primers and a probe were developed to explore the distribution and occurrence of the isolated diazotroph throughout the North Atlantic Ocean. Samples from a weekly time series in the Bedford Basin, bi-annual sampling on the Scotian Shelf and the GEOVIDE transect from Portugal to Labrador via Greenland were investigated (Figure 5.5). Abundances ranged from 0 – 2.5 x 10^5 *nifH* copies L^{-1} (highest abundances were found during the spring bloom in the Bedford Basin)

with a mean of $6.9 \times 10^3 \pm 5.9 \times 10^4$ *nifH* copies L^{-1}. On the Scotian Shelf, copy numbers averaged 134 ± 313 *nifH* copies L^{-1} without significant differences between the seasons ($p = 0.53$; Figure 5.5A and Figure 5.5B). The Bedford Basin showed significant seasonality with highest copy numbers in spring (mean of $1.3 \times 10^4 \pm 2.0 \times 10^4$ *nifH* copies L^{-1}, $p = 0.04$), followed by autumn, summer and winter, the latter without statistically significant differences (means of $7.6 \times 10^3 \pm 1.6 \times 10^4$, $4.3 \times 10^3 \pm 1.0 \times 10^4$ and $4.2 \times 10^3 \pm 6.2 \times 10^3$ *nifH* copies L^{-1} respectively, $p = 0.85$; Figure 5.5E). The GEOVIDE cruise showed two distinct areas of high abundance that were significantly different (Figure 5.5D). Samples east of 25°W and south of 54°N averaged $1.5 \times 10^4 \pm 3.5 \times 10^4$ *nifH* copies L^{-1}, while samples west of 25°W and north of 54°N had a mean *nifH* copy number of $1.1 \times 10^3 \pm 3.7 \times 10^3$ copies L^{-1} ($p < 0.01$, Figure 5.5D).

To further investigate the distribution of this novel diazotroph, a BLAST search was conducted and *nifH* sequences with 99% identity have been previously reported in the Chilean OMZ and in the deep North Pacific Ocean (Fernandez et al. 2011; Mehta et al. 2005; Altschul et al. 1990).

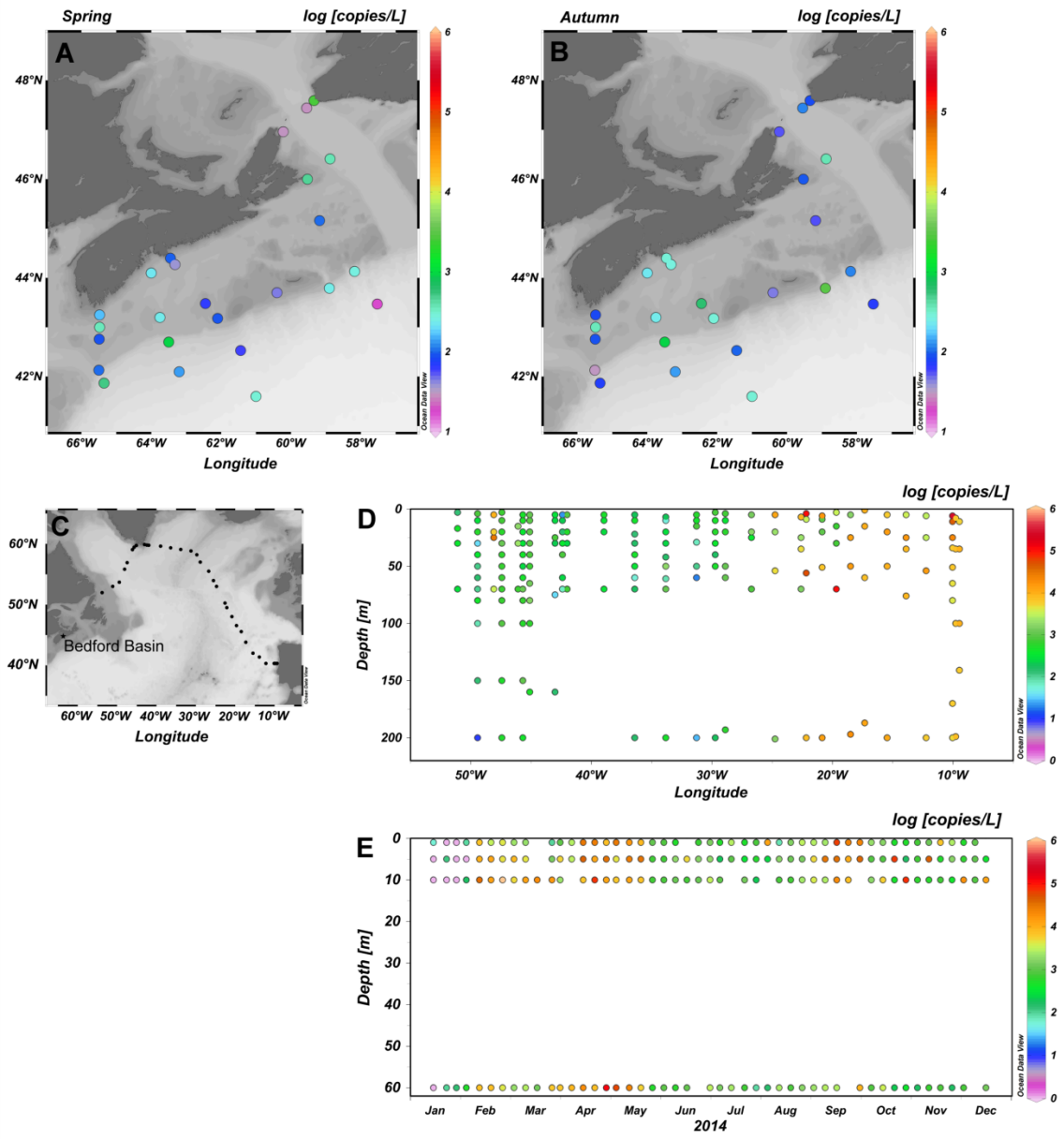


Figure 5.5: Distribution and abundances of the isolated diazotroph in the North Atlantic, assessed from an phylotype-specific TaqMan qPCR assay for the *nifH* gene.

Total *nifH* copies L^{-1} of the isolated diazotroph in the temperate North Atlantic as obtained by TaqMan qPCR. Abundances on the Scotian Shelf in autumn 2014 (A), and spring 2014 (B), along the GEOVIDE transit in summer 2014 (D) and throughout weekly sampling in the Bedford Basin in 2014 at 1, 5, 10 and 60 m depth (E) are depicted. The GEOVIDE cruise transect and location of the Bedford Basin are indicated on panel C. The maps were created in ocean data view (Schlitzer 2015)

Environmental parameters and *nifH* copy numbers were available for a total of 549 samples (Supplemental Table 10). To investigate environmental preferences of the isolated diazotroph, a principle component analysis (PCA) was conducted. The first three components explained 77.7% of variation in the data. Figure 5.6 shows the first two principle components with *nifH* abundances overlaid as circles. This analysis, combined with a BEST test, shows that this organism has no preference concerning DIN or silicate concentrations, temperature or depth. The most important factor accounting for abundance was salinity (sample statistic: 0.075), followed by salinity + O_2 (sample statistic: 0.074). Any further addition of environmental factors reduced the correlation significantly.

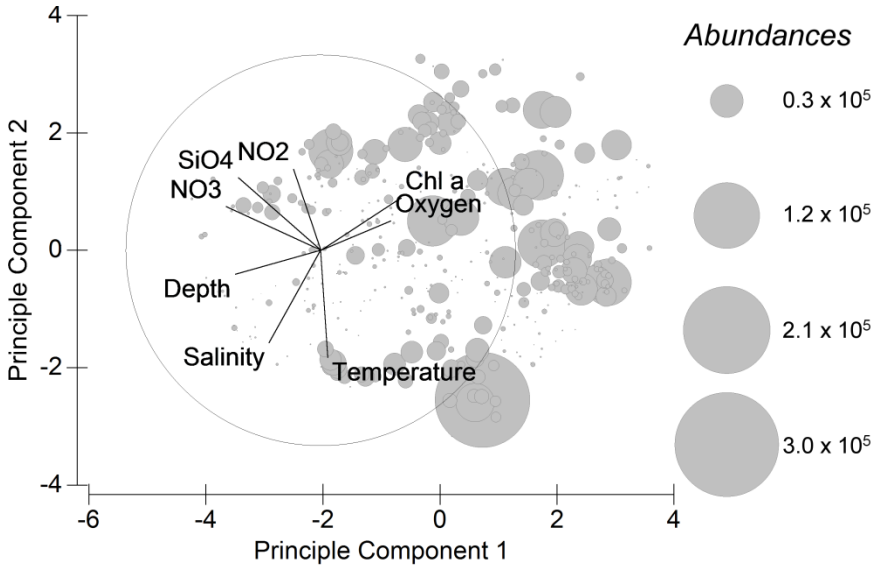


Figure 5.6: Environmental preferences of the isolated diazotroph.

Principal Components Analysis (PCA) of environmental parameters from the 2014 Bedford Basin sampling, autumn and spring cruises of the AZMP HUD2014004 and HUD2014030 (2014), and the 2014 GEOVIDE cruise. Abundances of the isolated diazotroph as obtained by qPCR are overlaid as circles.

5.4. Discussion

The diazotroph isolated in January 2014 from the Bedford Basin is a Gram-negative rod (Figure 5.1). The absence of carbon fixation genes indicates that the isolate is a heterotroph. Phylogenetic analysis of the full length 16S rRNA gene showed that it belongs to the gamma-Proteobacterial order Oceanospirillales (Figure 5.2A). The *nifH* gene aligns most closely with cluster I non-cyanobacterial *nifH* sequences (Figure 5.2B; Zehr et al. 2003).

Gamma-Proteobacteria are extremely diverse and they are the second most abundant phylum in the oceans (Sunagawa et al. 2015; Williams et al. 2010); it has been proposed that this also applies to the marine diazotrophs (Farnelid et

al. 2011). While few heterotrophic diazotrophs from this clade have been isolated (Bentzon-Tilia et al. 2014, 2015; Farnelid et al. 2014; Bostroem et al. 2007; Loveless 1999; Shieh et al. 1989; Maruyama et al. 1970), the few available isolates have already given a glimpse of the diversity of non-cyanobacterial diazotrophic genomes and metabolism. In addition to N_2 fixation, they most likely contribute to the bacterial community through many other functions, such as aromatic carbon degradation (Bentzon-Tilia et al. 2015).

5.4.1. Nutrient metabolism

The genome of the isolate suggests that it is very metabolically versatile; among 463 subsystems, the genome included pathways for nitrogen, iron, O_2 and carbon metabolism (4003 coding regions). The *nif* operon is present in two clusters, including the transcription regulator *nifA*, the three structural genes *nifHDK* and several synthesis support factors, as well as *nif* genes of unknown function (Figure 5.3A). There is a relative change of GC content between the *nif* operon and the surrounding DNA (1.6% drop), which might be an indicator for horizontal gene transfer of the *nif* operon, as has been reported for other diazotrophic organisms (Zehr et al. 2003). However, if the *nif* operon was obtained through horizontal gene transfer, it either occurred a long time ago or it did not occur between the isolate and *Marinobacterium litorale*, the reference organism with the most similar *nifH* sequence (85%), or *Azotobacter chroococcum*, which has the highest *nifHDK* similarity (80%), because their *nif* genes are contained in a single region instead of two, and are arranged in a

different order. The large differences between the isolated diazotrophs support the idea that diazotrophic reference genomes from the gamma-Proteobacterial clade are underrepresented in current data sets.

The presence of *nifH* transcripts during growth in nitrogen-depleted medium suggests the presence of the nitrogenase enzyme, which would confer the ability to perform N_2 fixation (Supplemental Figure 13). The genome includes several other ammonia assimilation pathways in addition to N_2 fixation (Figure 5.3B). Ammonia can be obtained by remineralization through the urea cycle, cyanate detoxification and nitrate uptake; the activity of these pathways is supported by increased growth of the isolate following nitrate addition (Figure 5.3B, Supplemental Figure 12). How the presence of these substrates regulates N_2 fixation rates in this organism and whether they may be connected in a specific ratio to the presence of carbon, phosphate, iron and O_2 species (as has been proposed for heterotrophic diazotrophs; Bentzon-Tilia et al. 2015; Fernendez et al. 2015; Rieman et al. 2010), remains to be evaluated.

The specific metabolic requirements for iron and low O_2 result from the high iron content of the nitrogenase and its susceptibility to oxidation in the presence of O_2. The nitrogenase contains at least twenty iron ions per enzyme, resulting in estimates of a twenty-fold higher iron requirement for cyanobacterial diazotrophs compared to non-diazotrophic cyanobacteria (Berman-Frank et al. 2007, 2001; Kustka et al. 2003). The isolate contains an array of iron regulation, iron (III) and heme uptake transporters and processing pathways, making it well suited to acquire iron in the iron-limited surface ocean . Additionally, the isolate has bi-

directional transporters for several other trace metals to maintain a balance of essential metals and to export potentially toxic metals (As, Cd, Co, Cr, Cu, Hg, Mg, Na, Ni, Se and Zn).

It is still not understood how heterotrophic diazotrophs maintain an anaerobic environment for their O_2-sensitive nitrogenase. The enzymatic reaction centre is destroyed upon oxidation: dependence on low O_2 concentrations may suggest that the enzyme evolved before the earth's atmosphere became oxygenated (Summons et al. 1999; Falkowski 1997). The two proposed mechanisms for O_2 evasion in non-cyanobacterial diazotrophs are metabolic protection, and limiting N_2 fixation to low O_2 environments. Multiple mechanisms of metabolic protection have been observed in terrestrial organisms that can not necessarily be observed on the genomic level, but possibly on the translational or observational level: protection from oxidative stress through capsule formation, increased respiration leading to O_2 drawdown, increased expression of reducing equivalents, conformational changes of the nitrogenase, and formation of a protein complex around the nitrogenase (Inomura et al. 2017; Schlesier et al. 2015; Paulus et al. 2012; Sabra et al. 2000; Poole and Hill 1997; Thorneley, Ashby 1989; Dingler et al. 1988; Dingler and Oelze 1987). When grown on nitrogen-depleted agar plates, the isolate forms shiny colonies, which can be indicative of capsule formation and it was observed to flocculate in DIN-depleted liquid cultures, a mechanism that has been shown to decrease intracellular O_2 by decreasing the exposed surface area and hence reducing O_2 diffusion (Bible et al. 2015). The genome did contain genes for several enzymes to deal with O_2 and reactive O_2 species, including

peroxidases (peroxide removal), glutaredoxins (general antioxidant) and the broader glutathione redox metabolism (regeneration of glutaredoxins), superoxide dismutase (O_2^- removal), catalase (H_2O_2 removal) and nitric oxide dioxygenase (NO and O_2 removal) as well as the direct removal of O_2 (rubredoxins and dioxygenases). These pathways are commonly found as general protection against oxidative damage, but upregulation under oxidative stress could contribute to maintaining low intracellular O_2 concentrations (Cabiscol et al. 2010).

The other proposed mechanism of O_2 evasion is the inhabitation of microanaerobic environments on respiratory active particles (Riemann et al. 2010). Although the detection of *nifH* sequences from the anaerobic cluster II throughout the open oceans supports this hypothesis, it is hard to observe this mechanism in pure cultures (Riemann et al. 2010).

The genome of the isolate suggests broad metabolic versatility in terms of carbon metabolism with a suite of possible carbon sources ranging from acetate, mono-, di- and polycarbohydates to aromatic carbon compounds (127 genes). The recently isolated diazotrophs *Pseudomonas stutzeri* BAL361, *Rhodopseudomonas palustris* BAL398 and *Rhodospirillum ornithinolytica* BAL286 (Bentzon-Tilia et al. 2015) also harbour a large number of genes related to carbon metabolism, including those for aromatic compound degradation. The most closely related described species to the isolate based on the 16S rRNA gene is *Thalassolituus oleivorans* (97.2% sequence similarity). *T. oleivorans* was isolated from oil-contaminated oceanic regions and belongs to the group of

obligate hydrocarbonoclastic marine bacteria (OHCB; Golyshin et al. 2013; Yakimov et al. 2004). There is no evidence for the common hydrocarbon degradation genes in the Bedford Basin isolate's genome (Alkane 1-monooxygenase encoded by $alkB_1$ or $alkB_2$), however, addition of Tween 20 (a molecule consisting of 20 repeat units of polyethylene glycol) to the growth medium increased growth rates of the isolate (Jennifer Tolman, personal communication).

Although it has been hypothesized that the distribution and N_2 fixing activity of cyanobacteria are dependent on dissolved P and fixed N concentrations, it is likely that the availability of organic carbon substrates is also a factor for heterotrophic diazotrophs in the natural environment (Fernandez et al. 2015; Deutsch et al. 2007). N_2 fixation and the reduction of intracellular O_2 is a highly energy-demanding process (at least 16 ATP per fixed N_2; Großkopf and LaRoche 2012; Postgate 1982), and heterotrophic diazotrophs rely on organic carbon sources for energy supply. The isolate's ability to process a wide variety of carbon compounds expands its feeding niche and could be advantageous in carbon limited environments. Additionally, the presence of the complete polyhydroxybutyrate metabolism pathways in the isolate's genome with the observation of inclusions at the cell poles in TEM images suggests the storage of carbon in the form of polyhydroxyalkanoates (PHA; Figure 5.1). PHA could provide the isolated heterotroph with an energetic carbon source when dissolved organic carbon is scarce in the environment. PHAs have recently gained major

interest in the biotechnology industry, because of their resemblance to plastic and their biodegradability (Reddy et al. 2003).

5.4.2. Distribution

We performed specific *nifH* qPCR on samples collected for 48 consecutive weeks in the Bedford Basin at 1, 5, 10 and 60 m depth in 2014, in spring and autumn of 2014 on the Scotian Shelf and along the GEOVIDE cruise transect from Portugal to Labrador via Greenland in the summer of 2014, amounting to a total of 549 samples (Figure 5.5).

Clustering samples according to their environmental and hydrographic parameters did not result in significant separation of the cruise transects (Supplemental Figure 14A). Instead, environmental conditions spanned wide ranges (e.g. temperature: -1.5 – 23.7°C, salinity: 27.0 – 36.4, NO_3^-: below detection limit - 24 µM, O_2: 48.7 – 738.7 µM). PCA analysis showed that environmental parameters divided the Bedford Basin samples into a high and low nutrient cluster (Supplemental Figure 14B). The low nutrient cluster overall resembled oceanic conditions more than the high nutrient cluster, though salinity was generally lower, which was expected since the Bedford Basin receives fresh water influx through the Sackville river and rain water runoff (mean = 30.08 ± 0.92, oceanic mean = 33.93 ± 1.70; Supplemental Figure 14B).

Copy numbers of the *nifH* gene ranged from 0 – 2.5 x 10^5 *nifH* copies L^{-1}, which is three times higher than maximum *nifH* abundances recorded for a widespread

gamma-proteobacterial diazotroph (8.0 x 10^4 *nifH* copies L^{-1}; Langlois et al. 2015). It falls within the range of maximum *nifH* copy numbers recorded for the intensely investigated cyanobacterium *Candidatus* Atelocyanobacterium thalassa (maximum *nifH* copy numbers of 2 x 10^4, 1.3 x 10^5 and 2.2 x 10^6 *nifH* copies L^{-1} in the North Atlantic Ocean; Turk et al. 2011; Moisander et al. 2010; Langlois et al. 2008) and are four-fold smaller than maximum abundances of *Trichodesmium spp.*, which has been thought to contribute the most significantly to N_2 fixation rates (1 x 10^9 *nifH* copies L^{-1}; Luo et al. 2012). However, it has been shown that *Trichodesmium spp.* is polyploid; multiple copies of the *nifH* gene per cell result in an overestimation of abundance by up to 40-fold (Sargent et al. 2016). Genome sequencing has demonstrated that there is only a single copy of *nifH* in the Bedford Basin isolate's genome, though this is not evidence of a single genome per cell. Whether copy numbers can be directly related to numbers of cells remains to be determined. However, the relative abundances of the isolate obtained from qPCR assays revealed a wide distribution and significant abundance throughout the North Atlantic Ocean in comparison to other diazotrophs. This suggests that the isolates is an important part of the diazotrophic bacterial community.

Furthermore, an NCBI database search using the *nifH* gene showed that the isolate has previously been found in the OMZ off the Chilean coast and in the deep North Pacific; the 16S rRNA gene has been identified in additional high-throughput sequencing data throughout the North Atlantic Ocean with more frequent detection at higher latitudes (personal data), indicating its presence

throughout the global oceans (Fernandez et al. 2011; Mehta et al. 2005; Altschul et al. 1990).

The *nifH* gene abundances for the isolate were highly variable throughout the cruises (Figure 5.5). The highest abundances were measured on the south-eastern cruise transect of the GEOVIDE cruise and during the spring period in the Bedford Basin (averages of 1.5×10^4 and 1.3×10^4 *nifH* copies L^{-1} respectively), although the environmental conditions were substantially different in those two sample subsets (Supplemental Figure 14C). The Bedford Basin experienced the spring bloom during this period (starting on March 19). The warmed surface waters of the Bedford Basin started to stratify and phytoplankton species bloomed as seen in chlorophyll a concentrations (Supplemental Table 10; Chapter 6). The increase in abundance of the isolate paralleled the phytoplankton bloom progression. A less prominent increase of *nifH* copies of the isolate was also observed during the autumn bloom (from September 17) when phytoplankton thrived as a result of water column mixing and the replenishment of nutrients to the nutrient-deplete surface (Supplemental Table 10; Chapter 6). An increase of heterotrophic *nifH* sequences following a phytoplankton bloom was also observed during mesocosm experiments and a time series study off the Chilean coast (Fernandez et al. 2015; Turk-Kobo et al. 2015). Subsequently, it was proposed that heterotrophic diazotrophs may rely on organic material from phototrophs for their growth. The south-eastern section of the GEOVIDE transect was influenced by Gulf Stream water masses as indicated by higher temperatures and higher salinity measurements (Supplemental Figure 15)

possibly indicating a preference for those conditions. However, statistical analysis showed that lower salinity and higher O_2 concentrations predicted a higher *nifH* gene count. Because the Bedford Basin was less saline, and most of the high abundances were measured in the Bedford Basin, correlation with low salinity probably resulted from the Bedford Basin samples. O_2 concentrations were lowest during winter in the Bedford Basin and on the Scotian Shelf; both subsets accounted for isolate abundances that were 100-10,000 fold lower than maximum abundances, which might indicate the isolate's preference for oxygenated water masses, still leaving the possibility for metabolic adaptation during N_2 fixation or in microanaerobic niches. It is important to note that this analysis did not include all factors that could influence the distribution of the isolate: trace metals and carbon sources have been proposed to be important drivers of diazotrophic distribution, but no measurements were available for these parameters (Fernandez et al. 2015; Berman-Frank et al. 2007; Kustka et al. 2003; Berman-Frank et al. 2001).

5.5. Conclusion

We isolated a heterotrophic diazotroph of the phylum gamma-proteobacteria with a 16S rRNA sequence 97.2% identical to members of the order Oceanospirillales (*Thalassolituus oleivorans* R6-14). Based on a taxon-specific *nifH* qPCR assay, the novel isolate is widespread and abundant throughout the temperate North Atlantic Ocean. Closely related *nifH* sequences have been reported in the Chilean OMZ and the deep Pacific Ocean (Fernandez et al. 2011; Mehta et al.

2005). The isolate's contribution to N_2 fixation remains unclear in the environment and needs to be further investigated in laboratory experiments. Its abundance showed no correlation with low DIN, but a possible preference for less saline, oxygenated water masses. The metabolism inferred from its genome sequence, revealed a wide variety of genes related to iron uptake and oxidative stress protection mechanisms. The genome harbours genes for several carbon degradation pathways, including short and long chain carbohydrates and aromatic compounds. It can store carbon as inclusion bodies of PHA. Its wide distribution at higher latitudes across the entire North Atlantic suggests that it is a proteobacterial diazotroph of ecologic importance in the marine microbial community and will continue to shed light on the metabolism of heterotrophic diazotrophs.

5.6. Acknowledgements

Jenni-Marie Ratten's work was supported by the NSERC CREATE Transatlantic Ocean System Science and Technology (TOSST) grant, the Stiftung für Kanadastudien (SKS), the Deutscher Akademischer Austausch Dienst (DAAD) and Julie LaRoche's NSERC Discovery research grant. Thanks to Mary Ann Trevors of the electron microscope facility for processing samples for Transmission Electron Microscopy and also thanks to Ping Li of the Scientific Imaging Suite in the Biology Department at Dalhousie University for the support with Transmission Electron Microscopy imaging.

5.7. References

Akaraonye, E., Keshavarz, T. and Roy, I., 2010. Production of polyhydroxyalkanoates: the future green materials of choice. Journal of Chemical Technology and Biotechnology, 85(6), pp.732-743.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

Anzaldi, L.L. and Skaar, E.P., 2010. Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. Infection and immunity, 78(12), pp.4977-4989.

Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stöckel, J., Liberton, M., Min, H., Sherman, L.A. and Pakrasi, H.B., 2011. Novel metabolic attributes of the genus Cyanothece, comprising a group of unicellular nitrogen-fixing cyanobacteria. MBio, 2(5), pp.e00214-11.

Bentzon-Tilia, M., Farnelid, H., Jürgens, K. and Riemann, L., 2014. Cultivation and isolation of N_2-fixing bacteria from suboxic waters in the Baltic Sea. FEMS microbiology ecology, 88(2), pp.358-371.

Bentzon-Tilia, M., Traving, S.J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J.L., Markager, S. and Riemann, L., 2015. Significant N_2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. The ISME journal, 9(2), pp.273-285.

Berman-Frank, I., Cullen, J.T., Shaked, Y., Sherrell, R.M. and Falkowski, P.G., 2001. Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. Limnology and Oceanography, 46(6), pp.1249-1260.

Berman-Frank, I., Quigg, A., Finkel, Z.V., Irwin, A.J. and Haramaty, L., 2007. Nitrogen-fixation strategies and Fe requirements in cyanobacteria. Limnology and Oceanography, 52(5), pp.2260-2269.

Bible, A.N., Khalsa-Moyers, G.K., Mukherjee, T., Green, C.S., Mishra, P., Purcell, A., Aksenova, A., Hurst, G.B. and Alexandre, G., 2015. Metabolic adaptations of Azospirillum brasilense to oxygen stress by cell-to-cell clumping and flocculation. Applied and environmental microbiology, 81(24), pp.8346-8357.

Bird, C., Martinez, J.M., O'Donnell, A.G. and Wyman, M., 2005. Spatial distribution and transcriptional activity of an uncultured clade of planktonic diazotrophic γ-Proteobacteria in the Arabian Sea. Applied and environmental microbiology, 71(4), pp.2079-2085.

Boetzer, M. and Pirovano, W., 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC bioinformatics, 15(1), p.211.

Bombar, D., Paerl, R.W. and Riemann, L., 2016. Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. Trends in microbiology, 24(11), pp.916-927.

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K.A., Moutin, T., Hamersley, R.M., Grosso, O., Zehr, J.P. and Capone, D.G., 2013. Aphotic N 2 fixation in the eastern tropical South Pacific Ocean. Plos one, 8(12), p.e81265.

Boström, K.H., Riemann, L., Kühl, M. and Hagström, Å., 2007. Isolation and gene quantification of heterotrophic N_2-fixing bacterioplankton in the Baltic Sea. Environmental microbiology, 9(1), pp.152-164.

Cabiscol, E., Tamarit, J. and Ros, J., 2010. Oxidative stress in bacteria and protein damage by reactive oxygen species. International Microbiology, 3(1), pp.3-8.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), pp.540-552.

Chen, Y.B., Zehr, J.P. and Mellon, M., 1996. Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* spp. IMS 101 in defined media: evidence for a circadian rhythm. Journal of Phycology, 32(6), pp.916-923.

Clarke, K.R. and Gorley, R.N., 2006. PRIMER version 6: user manual/tutorial. PRIMER-E, Plymouth, UK, 192.

Codispoti, L.A., 2007. An oceanic fixed nitrogen sink exceeding 400 Tg N a-1 vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences, 4(2), pp.233-253.

Deutsch, C., Sarmiento, J.L., Sigman, D.M., Gruber, N. and Dunne, J.P., 2007. Spatial coupling of nitrogen inputs and losses in the ocean. Nature, 445(7124), pp.163-167.

Dingler, C. and Oelze, J., 1987. Superoxide dismutase and catalase in Azotobacter vinelandii grown in continuous culture at different dissolved oxygen concentrations. Archives of microbiology, 147(3), pp.291-294.

Dingler, C., Kuhla, J., Wassink, H. and Oelze, J., 1988. Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations. Journal of bacteriology, 170(5), pp.2148-2152.

Doudna, J.A. and Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213), p.1258096.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), pp.1792-1797.

Falkowski, P.G., 1983. Enzymology of nitrogen assimilation. Nitrogen in the marine environment, pp.839-868.

Falkowski, P.G., 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO 2 in the ocean. Nature, 387(6630), pp.272-275.

Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å. and Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PLoS One, 6(4), p.e19223.

Farnelid, H., Bentzon-Tilia, M., Andersson, A.F., Bertilsson, S., Jost, G., Labrenz, M., Jürgens, K. and Riemann, L., 2013. Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. The ISME journal, 7(7), pp.1413-1423.

Farnelid, H., Harder, J., Bentzon-Tilia, M. and Riemann, L., 2014. Isolation of heterotrophic diazotrophic bacteria from estuarine surface waters. Environmental microbiology, 16(10), pp.3072-3082.

Fernandez, C., Farías, L. and Ulloa, O., 2011. Nitrogen fixation in denitrified marine waters. PLoS One, 6(6), p.e20539.

Fernandez, C., González, M.L., Muñoz, C., Molina, V. and Farias, L., 2015. Temporal and spatial variability of biological nitrogen fixation off the upwelling system of central Chile (35–38.5° S). Journal of Geophysical Research: Oceans, 120(5), pp.3330-3349.

Golyshin, P.N., Werner, J., Chernikova, T.N., Tran, H., Ferrer, M., Yakimov, M.M., Teeling, H., Golyshina, O.V. and MAMBA Scientific Consortium, 2013. Genome sequence of Thalassolituus oleivorans MIL-1 (DSM 14913T). Genome announcements, 1(2), pp.e00141-13.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M.C. and McCombie, W.R., 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome research, 25(11), pp.1750-1756.

Gregory, M.R., 2009. Environmental implications of plastic debris in marine settings—entanglement, ingestion, smothering, hangers-on, hitch-hiking and alien invasions. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364(1526), pp.2013-2025.

Großkopf, T. and LaRoche, J., 2012. Direct and indirect costs of dinitrogen fixation in Crocosphaera watsonii WH8501 and possible implications for the nitrogen cycle. Frontiers in microbiology, 3, p.236.

Guillard, R.R. and Ryther, J.H., 1962. Studies of marine planktonic diatoms: I. Cyclotella Nana Hustedt, and Detonula Confervacea (CLEVE) Gran. Canadian journal of microbiology, 8(2), pp.229-239.

Hamersley, M.R., Turk, K.A., Leinweber, A., Gruber, N., Zehr, J.P., Gunderson, T. and Capone, D.G., 2011. Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. Aquatic Microbial Ecology, 63(2), pp.193-205.

Haselkorn, R., 2007. Heterocyst differentiation and nitrogen fixation in cyanobacteria. In Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations (pp. 233-255). Springer Netherlands.

Hewson, I., Moisander, P.H., Achilles, K.M., Carlson, C.A., Jenkins, B.D., Mondragon, E.A., Morrison, A.E. and Zehr, J.P., 2007. Characteristics of diazotrophs in surface to abyssopelagic waters of the Sargasso Sea. Aquatic microbial ecology, 46(1), pp.15-30.

Imlay, J.A., Chin, S.M. and Linn, S., 1988. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. Science, 240(4852), p.640.

Inomura, K., Bragg, J. and Follows, M.J., 2017. A quantitative analysis of the direct and indirect costs of nitrogen fixation: a model based on Azotobacter vinelandii. The ISME Journal, 11(1), pp.166-175.

Jahson, S., Rai, A.N. and Bergman, B., 1995. Intracellular cyanobiont Richelia intracellularis: ultrastructure and immuno-localisation of phycoerythrin, nitrogenase, Rubisco and glutamine synthetase. Marine Biology, 124(1), pp.1-8.

Jayakumar, A., Al-Rshaidat, M.M., Ward, B.B. and Mulholland, M.R., 2012. Diversity, distribution, and expression of diazotroph nifH genes in oxygen-deficient waters of the Arabian Sea. FEMS microbiology ecology, 82(3), pp.597-606.

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D. and Stal, L., 2002. Dinitrogen fixation in the world's oceans. In The Nitrogen Cycle at Regional to Global Scales (pp. 47-98). Springer Netherlands.

Knapp, A.N., 2015. The sensitivity of marine N 2 fixation to dissolved inorganic nitrogen. The microbial nitrogen cycle, p.90.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv, p.071282.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution, p.msw054.

Kustka, A.B., Sanudo-Wilhelmy, S.A., Carpenter, E.J., Capone, D., Burns, J. and Sunda, W.G., 2003. Iron requirements for dinitrogen-and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations. Limnology and Oceanography, 48(5), pp.1869-1884.

Langlois, R.J., Hümmer, D. and LaRoche, J., 2008. Abundances and distributions of the dominant nifH phylotypes in the Northern Atlantic Ocean. Applied and environmental microbiology, 74(6), pp.1922-1931.

Langlois, R., Großkopf, T., Mills, M., Takeda, S. and LaRoche, J., 2015. Widespread distribution and expression of gamma A (UMB), an uncultured, diazotrophic, γ-proteobacterial nifH phylotype. PloS one, 10(6), p.e0128912.

Letunic, I. and Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, p.gkw290.

Loescher, C.R., Großkopf, T., Desai, F.D., Gill, D., Schunck, H., Croot, P.L., Schlosser, C., Neulinger, S.C., Pinnow, N., Lavik, G. and Kuypers, M.M., 2014. Facets of diazotrophy in the oxygen minimum zone waters off Peru. The ISME journal, 8(11), pp.2180-2192.

Loveless, T.M., Saah, J.R. and Bishop, P.E., 1999. Isolation of nitrogen-fixing bacteria containing molybdenum-independent nitrogenases from natural environments. Applied and environmental microbiology, 65(9), pp.4223-4226.

Madoui, M.A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P. and Aury, J.M., 2015. Genome assembly using Nanopore-guided long and error-free DNA reads. BMC genomics, 16(1), p.327.

Maruyama, Y., Taga, N. and Matsuda, O., 1970. Distribution of nitrogen-fixing bacteria in the central Pacific Ocean. Journal of the Oceanographical Society of Japan, 26(6), pp.360-366.

Mehta, M.P., Huber, J.A. and Baross, J.A., 2005. Incidence of novel and potentially archaeal nitrogenase genes in the deep Northeast Pacific Ocean. Environmental microbiology, 7(10), pp.1525-1534.

Mohr, W., Intermaggio, M.P. and LaRoche, J., 2010. Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium Crocosphaera watsonii WH8501. Environmental microbiology, 12(2), pp.412-421.

Moisander, P.H., Beinart, R.A., Hewson, I., White, A.E., Johnson, K.S., Carlson, C.A., Montoya, J.P. and Zehr, J.P., 2010. Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. Science, 327(5972), pp.1512-1514.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A. and Anson, E.L., 2000. A whole-genome assembly of Drosophila. Science, 287(5461), pp.2196-2204.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y. and Stepanauskas, R., 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. Journal of Computational Biology, 20(10), pp.714-737.

Orme-Johnson, W.H., 1985. Molecular basis of biological nitrogen fixation. Annual review of biophysics and biophysical chemistry, 14(1), pp.419-459.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R. and Fonstein, M., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research, 33(17), pp.5691-5702.

Parker, M.W., Pattus, F., Tucker, A.D. and Tsernoglou, D., 1989. Structure of the membrane-pore-forming fragment of colicin A.

Paulus, A., Rossius, S.G.H., Dijk, M. and de Vries, S., 2012. Oxoferryl-porphyrin radical catalytic intermediate in cytochrome bd oxidases protects cells from formation of reactive oxygen species. Journal of Biological Chemistry, 287(12), pp.8830-8838.

Piro, V.C., Faoro, H., Weiss, V.A., Steffens, M.B., Pedrosa, F.O., Souza, E.M. and Raittz, R.T., 2014. FGAP: an automated gap closing tool. BMC research notes, 7(1), p.371.

Postgate, J.R., 1982. Biology nitrogen fixation: fundamentals. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 296(1082), pp.375-385.

Reddy, K.J., Haskell, J.B., Sherman, D.M. and Sherman, L.A., 1993. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. Journal of Bacteriology, 175(5), pp.1284-1292.

Reddy, C.S.K., Ghai, R. and Kalia, V., 2003. Polyhydroxyalkanoates: an overview. Bioresource technology, 87(2), pp.137-146.

Riemann, L., Farnelid, H. and Steward, G.F., 2010. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. Aquatic Microbial Ecology, 61(3), pp.235-247.

Sabra, W., Zeng, A.P., Lünsdorf, H. and Deckwer, W.D., 2000. Effect of oxygen on formation and structure ofAzotobacter vinelandii alginate and its role in protecting nitrogenase. Applied and environmental microbiology, 66(9), pp.4037-4044.

Sahu, S.N., Acharya, S., Tuminaro, H., Patel, I., Dudley, K., LeClerc, J.E., Cebula, T.A. and Mukhopadhyay, S., 2003. The bacterial adaptive response gene, barA, encodes a novel conserved histidine kinase regulatory switch for adaptation and modulation of metabolism in Escherichia coli. Molecular and cellular biochemistry, 253(1), pp.167-177.

Sargent, E.C., Hitchcock, A., Johansson, S.A., Langlois, R., Moore, C.M., LaRoche, J., Poulton, A.J. and Bibby, T.S., 2016. Evidence for polyploidy in the globally important diazotroph *Trichodesmium*. FEMS Microbiology Letters, 363(21), p.fnw244.

Schlesier, J., Rohde, M., Gerhardt, S. and Einsle, O., 2015. A Conformational Switch Triggers Nitrogenase Protection from Oxygen Damage by Shethna Protein II (FeSII). Journal of the American Chemical Society, 138(1), pp.239-247.

Schlitzer, R., 2015. Ocean Data View. 2012.

Schneiker, S., dos Santos, V.A.M., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., Chernikova, T.N., Denaro, R., Ferrer, M., Gertler, C. and Goesmann, A., 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium Alcanivorax borkumensis. Nature biotechnology, 24(8), pp.997-1004.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), pp.2498-2504.

Shieh, W.Y., Simidu, U. and Maruyama, Y., 1989. Enumeration and characterization of nitrogen-fixing bacteria in an eelgrass (Zostera marina) bed. Microbial ecology, 18(3), pp.249-259.

Sohm, J.A., Subramaniam, A., Gunderson, T.E., Carpenter, E.J. and Capone, D.G., 2011. Nitrogen fixation by *Trichodesmium* spp. and unicellular diazotrophs in the North Pacific Subtropical Gyre. Journal of Geophysical Research: Biogeosciences, 116(G3).

Summons, R.E., Jahnke, L.L., Hope, J.M. and Logan, G.A., 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature, 400(6744), pp.554-557.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. and Cornejo-Castillo, F.M., 2015. Structure and function of the global ocean microbiome. Science, 348(6237), p.1261359.

Swannell, R.P., Lee, K. and McDonagh, M., 1996. Field evaluations of marine oil spill bioremediation. Microbiological Reviews, 60(2), pp.342-365.

Tamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution, 24(8), pp.1596-1599.

Tay, K.L., Doe, K.G., Wade, S.J., Vaughan, D.A., Berrigan, R.E. and Moore, M.J., 1992. Sediment bioassessment in Halifax harbour. Environmental toxicology and chemistry, 11(11), pp.1567-1581.

Thompson, A., Carter, B.J., Turk-Kubo, K., Malfatti, F., Azam, F. and Zehr, J.P., 2014. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. Environmental microbiology, 16(10), pp.3238-3249.

Thompson, R.C., Moore, C.J., Vom Saal, F.S. and Swan, S.H., 2009. Plastics, the environment and human health: current consensus and future trends. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1526), pp.2153-2166.

Thorneley, R.N. and Ashby, G.A., 1989. Oxidation of nitrogenase iron protein by dioxygen without inactivation could contribute to high respiration rates of Azotobacter species and facilitate nitrogen fixation in other aerobic environments. Biochemical Journal, 261(1), pp.181-187.

Tritt, A., Eisen, J.A., Facciotti, M.T. and Darling, A.E., 2012. An integrated pipeline for de novo assembly of microbial genomes. PloS one, 7(9), p.e42304.

Turk, K.A., Rees, A.P., Zehr, J.P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E.M.S. and Gilbert, J., 2011. Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. The ISME journal, 5(7), pp.1201-1212.

Turk-Kubo, K.A., Frank, I.E., Hogan, M.E., Desnues, A., Bonnet, S. and Zehr, J.P., 2015. Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon). Biogeosciences, 12(24), pp.7435-7452.

Turk-Kubo, K.A., Karamchandani, M., Capone, D.G. and Zehr, J.P., 2014. The paradox of marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. Environmental microbiology, 16(10), pp.3095-3114.

Van Beilen, J.B., Marin, M.M., Smits, T.H., Röthlisberger, M., Franchini, A.G., Witholt, B. and Rojo, F., 2004. Characterization of two alkane hydroxylase genes from the marine hydrocarbonoclastic bacterium Alcanivorax borkumensis. Environmental microbiology, 6(3), pp.264-273.

Vojvodic, A., Medford, A.J., Studt, F., Abild-Pedersen, F., Khan, T.S., Bligaard, T. and Nørskov, J.K., 2014. Exploring the limits: A low-pressure, low-temperature Haber–Bosch process. Chemical Physics Letters, 598, pp.108-112.

Voss, M., Croot, P., Lochte, K., Mills, M. and Peeken, I., 2004. Patterns of nitrogen fixation along 10 N in the tropical Atlantic. Geophysical Research Letters, 31(23).

Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J. and Birol, I., 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. GigaScience, 4(1), p.35.

Williams, K.P., Gillespie, J.J., Sobral, B.W., Nordberg, E.K., Snyder, E.E., Shallom, J.M. and Dickerman, A.W., 2010. Phylogeny of gammaproteobacteria. Journal of bacteriology, 192(9), pp.2305-2314.

Yakimov, M.M., Giuliano, L., Denaro, R., Crisafi, E., Chernikova, T.N., Abraham, W.R., Luensdorf, H., Timmis, K.N. and Golyshin, P.N., 2004. Thalassolituus oleivorans gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. International Journal of Systematic and Evolutionary Microbiology, 54(1), pp.141-148.

Yakimov, M.M., Timmis, K.N. and Golyshin, P.N., 2007. Obligate oil-degrading marine bacteria. Current opinion in biotechnology, 18(3), pp.257-266.

Zehr, J.P., 2011. Nitrogen fixation by marine cyanobacteria. Trends in microbiology, 19(4), pp.162-173.

Zehr, J.P., Jenkins, B.D., Short, S.M. and Steward, G.F., 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental microbiology, 5(7), pp.539-554.

Zehr, J.P., Mellon, M.T. and Zani, S., 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes. Applied and environmental microbiology, 64(9), pp.3444-3450.

Zehr, J.P., Bench, S.R., Carter, B.J., Hewson, I., Niazi, F., Shi, T., Tripp, H.J. and Affourtit, J.P., 2008. Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. Science, 322(5904), pp.1110-1112.

CHAPTER 6: ANNUAL CYCLE OF CHANGE IN BACTERIAL COMMUNITY STRUCTURE IN A TEMPERATE COASTAL MARINE EMBAYMENT IN THE NORTH ATLANTIC

Jenni-Marie Ratten, Jackie Zorz, and Julie LaRoche

Contribution of authors:

Jenni-Marie Ratten:	Sample extraction and processing, data analysis and drafting of manuscript
Jackie Zorz:	Technical support with sample extraction and collection
Julie LaRoche	Planning and discussion of manuscript

6.0. Abstract

Climate change and anthropogenic factors are altering both the marine environment and the composition of the marine microbial communities that are the major drivers of marine nutrient recycling. Establishing the current baseline of microbial community structure and continuing long-term time series observations are essential to predict how these microbially-controlled nutrient cycles may change in the future. The Bedford Basin is a coastal ocean-inlet on the Canadian east coast that is suitable for the establishment of such a time-series study.

Bacterial community structure was assessed at depths of 1, 5, 10, and 60 m at weekly intervals between January 15th 2014 and December 17th, 2014. Results from high-throughput tag sequencing for the 16S rRNA gene (V6-V8) included 7,744,556 reads that passed quality control, with a mean of 31,740 reads per sample, and a total of 23,056 Operational Taxonomic Units (OTU) identified in the entire dataset. Diazotrophic community diversity was also investigated using high throughput sequencing of the *nifH* gene on four dates (March 19th, June 18th, September 24th and December 17th) which resulted in 162,148 quality-controlled reads and 1122 *nifH* OTUs.

The dominant bacterial community in the surface (1, 5 and 10 m) transitioned from a diverse cold water community dominated by Rhodobacteraceae to a stratified warm water community displaying taxa such as the most-commonly detected Pelagibacteraceae, which are adapted to nutrient-depleted waters . During the phytoplankton spring bloom, the community was dominated by *Pseudoalteromonas*, *Psychrobacter* and *Ulvibacter* OTUs, whereas during the autumn bloom, specific Rhodobacteraceae were seen. The main drivers of community transition in the surface were identified to be temperature, nutrient, O_2 and chlorophyll concentrations. At depth, similar conditions were associated with community composition, although chlorophyll played a less important role, and *SUP05*, previously identified to be present in dysoxic water masses (20 – 90 μM O_2) was abundant at certain times of the year. Significant abrupt disruptions of the otherwise transitioning community occurred twice in the form of oceanic water intrusions.

The diazotrophic community composition was highly variable between seasons and depth, with exceptionally high diversity in winter surface waters. As seen in other temperate marine environments, *nifH* sub-cluster Ip sequences dominated throughout all samples, with cyanobacterial *nifH* sequences detected only in June and September surface samples. With a single exception, all *nifH* cyanobacterial sequences aligned with the *Candidatus* Atelocyanobacterium clade.

We were able to show that the bacterial community adapts rapidly to changes in the environment and that the diazotrophic community changed compositions with seasons, which supports previous findings that microbial communities will readily adapt to a changing ocean.

6.1. Introduction

Oceanic microbial communities are extremely diverse and dynamic. Their composition is tightly linked to both the physical environment and members of the local biological system. Changes in the environment such as shifting temperature, light intensity, and/or nutrient and O_2 concentrations significantly impact the microbial community composition (Giovannoni and Vergin 2012; Wright et al. 2012; Diaz and Rosenberg 2008; Arrigo 2005; Price and Sowers 2004; Pomeroy and Wiebe 2001). While marine microbes respond to external changes, they also play a crucial role in shaping their environment and community. Biological activity is responsible for the marine cycling of nutrients and elements, and in so doing, contributes to global element cycles (Zehr and Kudela 2011; Falkowski et al. 2008; Kirchman 2000).

For example, the biological pump provides a continuous sink for anthropogenic carbon through the fixation of CO_2 by primary producers and the eventual export of particulate organic matter into the deep sea. Diazotrophs provide newly fixed nitrogen to the system, counteracting losses through the anaerobic pathways of denitrification and anammox (Gruber 2008; Codispoti 2007; Hamersley et al. 2007; Codispoti et al. 2001; Azam 1998; Ingall et al. 1994). Aside from physical factors, microbial community composition is also controlled by biotic interactions; predator-prey interactions, viral infection, competition through uptake mechanisms, symbiosis, and allelopathy all determine the success or failure of one species over another (Strom 2008; Suttle 2007; Calbet and Landry 2004).

Establishing the effect of climate change and other anthropogenic impacts on the ocean biota has been a major driver of marine ecology and ecosystem research over the past decades (Doney et al. 2012). The response of microbial communities must be considered when looking at the effects and feedback mechanisms of short term and long term variability. Recent advances in molecular technologies and the drastic decrease in costs of next generation sequencing have enabled scientists to explore the composition and changes of marine microbial communities using culture-independent methods. These studies capture a vast amount of organisms that were traditionally not investigated, because of their resistance to cultivation. Despite increased efforts, most ocean areas remain drastically undersampled with respect to the distribution of microbes in time and space, often limited to sparse broad-scale transects across

oceanic basins without the possibility of repeat sampling within a period relevant to variation in microbial community structure (Fuhrman et al. 2015).

Time-series studies in temperate regions have been valuable in the study of changing oceans, as seasonality shows transitions that occur in microbial communities when well mixed water masses become stratified (Giovannoni and Vergin 2012). Recent time-series studies show that microbial communities exhibit characteristic shifts and cycles in the structure of abundant taxa over the seasons, suggesting that microbial communities will be significantly affected by increasing stratification (El-Swais et al. 2014; Karl and Church 2014; Gilbert et al. 2012; Giovannoni and Vergin 2012; Fuhrman et al. 2006). Through subsequent years of observation, time-series studies establish a baseline of microbial community structure as a normative reference from which deviations can be observed (Fuhrman et al. 2015; Karl and Church 2014; Giovannoni and Vergin 2012). Traditionally, monitoring occurs at monthly or quarterly intervals, which may be insufficient to capture important transitions and disruptions to the community such as bloom periods or severe weather (Karl and Church 2014; Gilbert et al. 2012; Giovannoni and Vergin 2012; Fuhrman et al. 2006). Observations obtained from a multiple sampling depth are desirable to explore shifts of the environment throughout the water column (Giovannoni and Vergin 2012).

Here we present an analysis of the microbial community composition in a temperate coastal North Atlantic Ocean inlet (Bedford Basin, Halifax, Canada) sampled weekly at four depths over the course of one year. Coastal regions of

the world's oceans are responsible for an estimated 19% of oceanic net primary productivity and are thus regions of intense economic and environmental interest (Field et al. 1998). Since 1991, the Bedford Institute of Oceanography (BIO) has undertaken weekly monitoring of factors affecting the plankton ecosystem at the Compass Buoy station in the Bedford Basin (44° 41' 37" N, 63° 38' 25" W; Li et al. 2008; Li and Dickie 2001). We extended these efforts by determining bacterial community structure at four depths (1, 5, 10 and 60 m) at the Compass Buoy station on a weekly basis using next-generation high-throughput sequencing of the 16S rRNA gene. Our high frequency time- and depth-resolved measurements provide a detailed picture of the microbial community structure progression throughout the seasons, demonstrating the ephemeral nature of bloom and bust within the bacterial community, and capturing the depth distribution and seasonal transitions of dominant bacterial taxa. We also monitored the diazotrophic community quarterly to observe the seasonal dynamics of this specialized community in the Bedford Basin.

6.2. Methods

6.2.1. DNA Sample Collection

In collaboration with Bedford Basin Monitoring Program of the Bedford Institute of Oceanography (BIO), water samples were collected at the Compass Buoy station in the Bedford Basin (44^0 41' 30'' N, 63^0 38' 30''W) on a weekly basis from January 15 to December 17, 2014. 500 mL of water from four depths (1, 5, 10,

and 60 m) were filtered onto 0.2 µm polycarbonate filters using vacuum filtration (max 5 mmHg). Filters were flash frozen in liquid nitrogen and stored at -80°C.

6.2.2. DNA extraction

DNA was extracted using the QIAGEN DNeasy Plant Mini Kit with a slightly modified protocol as follows. Filters were incubated with 50 µL of lysozyme solution (5 mg mL^{-1} in TE buffer) at room temperature for 5 min. 45 µL of proteinase K solution (20 mg mL^{-1} in PCR grade water) and 400 µL of AP1 lysis buffer from the QIAGEN DNeasy Plant Mini Kit were added, followed by a one hour incubation at 52°C and 225 rpm on an orbital shaker. RNA was digested with 4 µL RNaseA (QIAGEN) at 65°C. Then, all liquid was placed on a Qiashredder column and extraction continued according to the manufacturer's protocol with a final elution volume of 50 µL. Concentrations and purity were determined using the NanoDrop 2000 and samples were stored in aliquots at -80°C.

6.2.3. 16S rDNA and **nifH** *gene amplification and MiSeq library preparation*

Sequencing of the V6-V8 region of the bacterial 16S rRNA and *nifH* gene was performed using an Illumina MiSeq platform at the Integrated Microbiome Resource (IMR) of the Centre for Comparative and Evolutionary Biology (CGEB) at Dalhousie University (Halifax, Canada). Sequences were amplified with

custom 16S fusion primers containing universal primer sequences (B969F and BA1406R; Comeau et al. 2011), along with Illumina adapters and barcodes for multiplexing (Comeau et al. 2017). Amplifications were performed using two different dilutions of template (undiluted and 1:10) to prevent bias. 25 µL reactions contained: 5 µL of 5xHF PCR Buffer, 0.5 µL dNTPs (40 mM), 5 µL forward and 5 µL reverse primer (1 µM), 0.25 µL Phusion polymerase (2 U μL^{-1}; Thermo Scientific), 2 µL sample or negative control and 7.25 µL PCR-grade water. Cycling conditions were: initial denaturation at 98°C for 30 s, followed by 30 cycles of 10 s at 98°C, 30 s at 55°C and 30 s at 72°C, and a final extension of 4.5 min at 72 °C. PCR product quality was verified using the E-gel 96-well high-throughput system (Invitrogen).

The first amplification of the nested *nifH* PCR was completed in 25 µL reactions made up of 2.5 µL 10x buffer (Qiagen), 2 µL dNTPs (10 µM; Invitrogen), 4 µL $MgCl_2$ (25 mM, Qiagen), 2 µL each of *nifH* 3 and 4 primers (10 µM; IDT), 0.3 µL BSA (20 mg mL^{-1}; NEB), 2.5 µL template, 0.125 µL HotStar Taq polymerase (0.625 U; Qiagen) and 9.575 µl PCR grade water. Cycling conditions were 95°C for 15 min followed by 35 cycles of 95°C (60 s), 45°C (60 s), and 72°C (60 s), with a final 10 min at 72°C. The second PCR was performed like the first, but with a final reaction volume of 10 µL, a reduced $MgCl_2$ concentration (1.2 µL of 25 mM stock), *nifH* 1/2 primers, and 1 µL of template from the first reaction. PCR conditions were modified to an annealing temperature of 54°C and 28 PCR cycles.

The first PCR step was repeated at a 1:10 template dilution for *nifH* positive samples (March: 5 and 60 m, June: a, 10 and 60 m, September: 10 and 60 m, December: 1, 10 and 60 m). For sequencing on the Illumina MiSeq platform, PCR products from both first round amplifications were combined and purified using a GeneJet PCR purification kit (Thermo Scientific). The second round of amplification was repeated with custom fusion primers linking the *nifH* 1/2 primer sequence, the Illumina adaptor, and unique barcode sequences for multiplexing (Supplemental Table 11). Amplification was carried out in 25 µL reactions with reagent concentrations as above and a modified annealing temperature of 52°C.

Subsequently, all amplification products were cleaned and normalized using the SequalPrep Normalization Plate Kit (Invitrogen). Samples were then combined at equal volumes, quantified with the Qubit (Invitrogen), and loaded into the Illumina MiSeq platform as a 20 pM final denatured library according to manufacturer's instructions. In total, 196 weekly samples of 1, 5, 10 and 60 m from January 15th 2014 till December 17th 2014 were sequenced for 16S rRNA and 10 samples returned *nifH* sequences (March 19th – 5 and 60 m, June 18th – 1, 10 and 60 m, September 24th – 10 and 60 m, and December 17th – 1, 10 and 60 m).

6.2.4. Bioinformatic analysis

Raw Illumina paired-end reads of the 16S rRNA V6-V8 regions and *nifH* were preprocessed for QIIME (Quantitative Insights Into Microbial Ecology; Caporaso et al. 2010a) using the 16S amplicon analysis flow of the IMR (https://github.com/mlangill/microbiome_helper#16s-workflow-starting-with-

demultiplex-miseq-fastq-files; Comeau et al. 2017). Initial steps were run for both analyses. Reads were stitched together using PEAR (Paired-End reAd merger; Zhang et al. 2014). Short (16S rRNA: <400bp; *nifH* < 330 bp), low quality (Q<30 in >10% of samples; Langille, https://github.com/mlangill), chimeric (using UCHIME; Edgar et al. 2011), or N containing reads were removed using Comeau et al.'s (2017) pipeline. The splitting of 16S from *nifH* sequences was achieved through an in-house script using *nifH* primer sequences as the determining factor. Remaining reads were fed into the QIIME pipeline (Caporaso et al. 2010a). The QIIME open reference picking pipeline was run using SortMeRNA for reference picking against the Greengenes database or a curated *nifH* database and sumaclust for de novo OTU picking (16S rRNA at 97 % identity, *nifH* at 96% identity; Mercier et al. 2013; Kopylova et al. 2012; McDonald et al. 2012; Werner et al. 2012). PyNAST was used to perform the alignments (Caporaso et al. 2010b). The last de-novo picking step, included in the pipeline by default, was suppressed due to an extremely long processing time caused by the large amount of reads and the subsampling percentage was changed from 0.1 % to 1.0 %. Subsequently, singletons and low confidence OTUs were removed from the dataset. For 16S rRNA genes, taxonomies were assigned with RDP Classifier 2.2 (Wang et al. 2007). Additional quality control measures included removing all sequences belonging to archaea and chloroplasts from the dataset by applying the QIIME *filter_fasta.py* script. Samples were rarefied to 9900 reads per sample for the 16S rRNA dataset and 1500 reads per sample for the *nifH* dataset, the lowest number of reads present in a single sample.

6.2.5. Environmental parameters

Environmental measurements from the CTD were performed by BIO and the CERC.OCEAN group at Dalhousie University in conjunction with the collection of water samples for DNA and flow cytometry analysis (Li 2014). For statistical analysis, the environmental matrix was processed in PRIMER-E V 6.1.12 (Clarke and Gorley 2006). Samples with missing values were deleted from the dataset, resulting in a matrix with 181 samples. After a correlation matrix was generated (Draftsman Plot), environmental variables that were correlated >0.8 were treated as one. This matrix was first log-transformed and then normalized in order to bring all variables to the same orders of magnitude.

6.2.6. Statistical Analyses

Statistical analyses were performed in R, QIIME and PRIMER-E version 6.1.12 (R Core Team 2015; Caporaso et al. 2010a; Clarke and Gorley 2006). Plots were generated in R using *ggplot2* or Ocean Data View version 4.6.5 unless stated otherwise (Schlitzer 2015; Wickham and Chang 2015). The rarefaction curve and diversity measures were calculated in QIIME (Caporaso et al. 2010a). Heatmaps were created in R using *gplots* and *RColorBrewer* (Warnes 2016; R Core Team 2015; Neuwirth 2011). Community matrices were Hellinger transformed and Bray-Curtis similarities were generated using the R package *vegan* (Dixon 2003). Also using the package *vegan*, Bray-Curtis similarities were used to depict patterns of community composition over the year as well as in two-dimensional space using nonmetric multidimensional scaling (NMDS; Dixon 2003).

Community interactions were determined using SparCC, an algorithm that can identify community interactions in high-throughput sequencing data more accurately than the more simplistic approaches of Pearson or Spearman correlation (Friedman and Alm 2012). For analysis in Primer–E version 6.1.12, the abundance matrix was limited to the 100 most abundant OTUs and divided into surface (1, 5 and 10 m) and deep (60 m) samples, as these groups exhibited differing general patterns in environmental parameters (Clarke and Gorley 2006). The following statistical tests were carried out: a BEST (Bio-Env + Stepwise) test identified those environmental factors that best explain community composition. The comparison was carried out between the transformed and normalized environmental matrix and the Hellinger transformed Bray–Curtis similarities of the OTU table. A combination of variables, which showed the maximum correlation with OTU occurrence, was identified. These variables each were divided into two groups of smaller and greater value, determined by LINKTREE analysis, which finds the best thresholds based on OTU abundances. An ANOSIM test was performed to determine whether OTUs were significantly distributed between predefined groups. For each significant value, the discriminating phylotypes of the 24 most abundant (making up more than 1% of OTUS over the year) OTUs were identified using the SIMPER routine. Lastly, the time lag of the 24 most commonly occurring OTUs between surface (1, 5 and 10 m) and deep (60 m) samples was determined using the cross correlation function of the Applied Statistical Time Series Analysis (*astsa*) package in R (Stoffer 2012). This package is designed to find correlations between two time series datasets.

6.2.7. Phylogenetic analysis

The *nifH* OTUs isolated from the Bedford Basin were aligned based on their protein sequence using MAFFT v. 7, returned to nucleotide sequences with PAL2NAL, ambiguous sequence alignment regions were removed using Gblocks and a maximum likelihood tree was inferred using RAxML with the GTR-GAMMA model using rapid bootstrapping with 100 replicates. Default parameters were applied (Yamada et al. 2016; Stamakatis 2014; Suyama et al. 2006; Katoh et al. 2002; Castresana 2000). The branch lengths of the tree showed the number of substitutions per site. The tree was depicted using iTOL (interactive Tree of Life; Letunic and Borg 2016).

6.3. Results

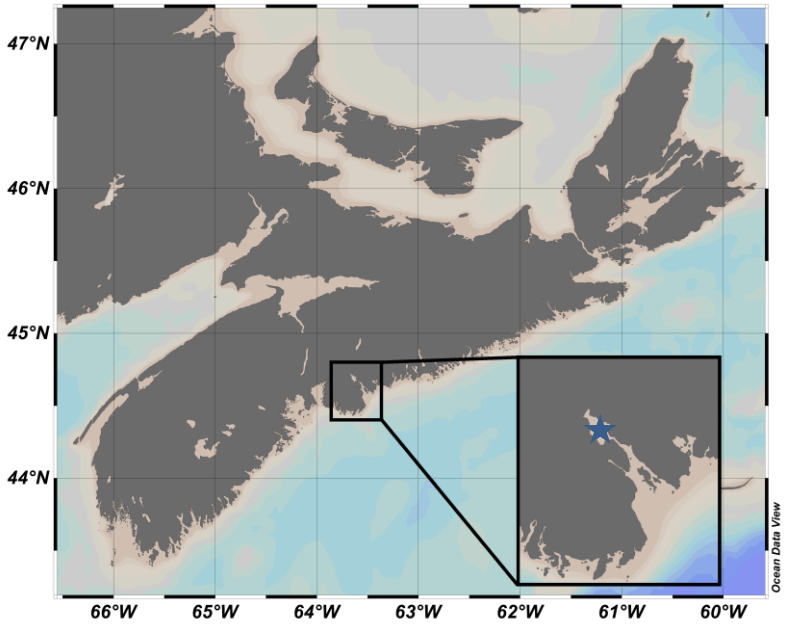


Figure 6.1: Map of the sampling site.
The Bedford Basin is located on the eastern coast of Nova Scotia, Canada at 44.69°N and 63.64°W (blue star).

6.3.1. Time-series site description

The Bedford Basin, located to the north-west of the Halifax Peninsula, is connected to the open Atlantic Ocean by a 10 km long, 400 m wide and 20 m deep channel (Li and Dickie 2001) (Figure 6.1). Freshwater is supplied by the Sackville River and by runoff. Since 1991 the Bedford Institute of Oceanography (BIO) has carried out weekly monitoring at the Compass Buoy station, located at the deepest point of the basin (71 m); water samples are collected at depths of 1, 5, 10 and 60 m. BIO and the CERC.OCEAN group at Dalhousie University monitor physical, chemical, biological and optical parameters, including water

temperature, nutrient concentration, phytoplankton abundance and light attenuation. Using CTD data (personal communication Richard Davis, CERC.OCEAN Dalhousie University), we have depicted relevant environmental conditions throughout 2014 in Figure 6.2. The well-mixed, nutrient-rich waters in the cold late autumn and winter months showed suboxic conditions at depth. In the spring, increases in irradiance and temperature led to water column stratification and spring bloom initiation at week 15. Chlorophyll concentrations decreased when nutrients became depleted in the surface layer during the summer months. The autumn bloom started at week 32 and peaked at week 38, at least for the bacterial community (Figure 6.2). Surface water temperatures increased the most in the summer, whereas bottom waters remained cold throughout the year.

On two occasions, abrupt changes occurred in the physical environment of the Bedford Basin, most prominently seen at 60 m. During weeks 10 to 11 and weeks 28 to 29, drastic shifts are seen in nutrient concentrations as well as salinity due to intrusion of offshore waters. Both incidents are followed by increases in chlorophyll concentration and overall fluorescence in the 1, 5 and 10 m samples. Bacterial counts as obtained by flow cytometry steadily increased over the summer in the surface until it decreased rapidly at the start of the autumn bloom (1.4×10^3 cells μL^{-1} in week 20 to a recorded maximum of 1.2×10^4 cells μL^{-1} in week 37 with a yearly mean of $2.5 \times 10^3 \pm 2.2 \times 10^3$ cells μL^{-1}). Counts at depth ranged from 111 to 5.2×10^3 cells μL^{-1} (mean = $1.6 \times 10^3 \pm 755$ cells μL^{-1}).

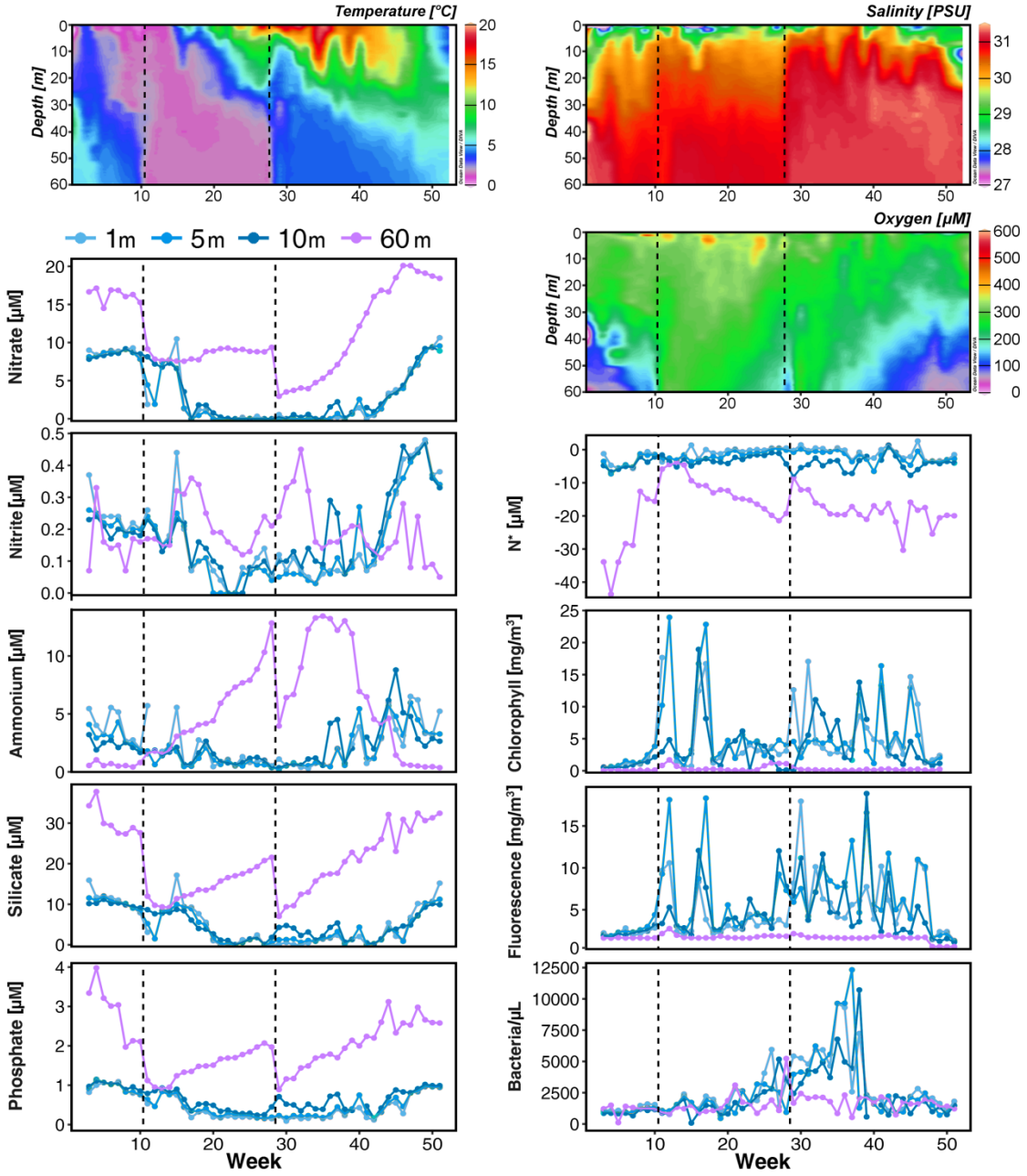


Figure 6.2: Environmental conditions in the Bedford Basin in 2014.

Environmental conditions are displayed as obtained from CTD measurements by BIO during weekly sampling at the deepest point of the basin (bottom depth: 71 m). The four sampling depths (1, 5, 10 and 60 m) are colored as indicated by the legend. Disruptions of the water column due to deep water intrusions occurred during weeks 10/11 and 28/29, represented as dotted lines.

6.3.2. Overview of taxa

In total, 196 samples were collected for DNA extraction between January 15th, 2014 and December 17th, 2014. PCR amplifications of variable regions V6-V8 from 16S rDNA were successful, with the exception of one sample removed from analysis due to low read numbers (October 1st, 10 m). Altogether 7,744,556 sequences passed quality control with a mean of 31,740 reads per sample. Operational taxonomic units (OTU) were clustered with a threshold of 97% sequence identity, which classified 23,056 OTUs. Samples were rarefied to 9,900 sequences per sample. Alpha-rarefaction analysis showed that most samples had not reached complete saturation at this sequencing depth; however, diversity analysis showed that higher rarefaction did not change overall diversity measures, suggesting that all major OTUs are included at 9,900 sequences per sample (Supplemental Figure 16).

A total of 25 taxa making up more than 1% of reads for each depth combined are depicted in Supplemental Figure 17 and listed in Supplemental Table 13 including their GI number and prokMSA ID. This included 17 OTUs at 1 m, 18 OTUs at both 5 m and 10 m and 11 OTUs at 60 m depth. Within the surface samples (1, 5 and 10 m), 11 of these highly abundant OTUs were identical. The three most abundant OTUs in the surface (1, 5 and 10 m, respectively) were RhodobacteraceaeA (9.4%, 7.0% and 6.2%, respectively), *Octadecabacter* (11.8%, 6.9% and 6.0%, respectively) and RhodobacteraceaeB (11.6%, 6.2% and 5.4%, respectively). The > 1% OTU community contained 4 OTUs that were found only at depth and 9 OTUs that are absent from the deep water but present

throughout all surface depths. The three most common OTUs at depth were PelagibacteraceaeA (12.5%), *Octadecabacter* (4.1%) and RhodobacteraceaeA (3.7%). Overall, the community of OTUs that made up > 1% of reads contributed between 16% (week 3, 1 m) and 83% (week 12, 10 m) to all sequences in the samples.

6.3.3. Community Diversity

The alpha diversity measures Shannon diversity and Chao1 were calculated for each month and depth for 2014 (Figure 6.3). The observed Shannon diversity varied over the seasons and little between surface and depth (Figure 6.3A). Diversity reached its peak after the collapse of the autumn bloom in November (week 43) and remained at these levels until February (highest values were 6.9, 7.4 and 7.2 for 1, 5 and 10 m respectively; Figure 6.3). Overall bacterial counts were lowest after the autumn bloom and throughout the winter period (Figure 6.2). The spring bloom (week 16) showed the lowest diversity values throughout the year (2.9, 2.6 and 1.5 for the surface depths, Figure 6.3). Diversity also decreased prior to the peak of the autumn bloom (5.1, 5.4 and 5.5, weeks 34 – 38, Figure 6.3), but not as drastically as during the spring bloom. Variance in diversity was highest during the spring bloom at all depths and relatively low throughout the rest of the year. At 60 m, highest diversity was detected in February (7.5, week 9), with lowest diversity and highest variance during the spring bloom (2.0, week 16). The Shannon diversity index indicated that the

bacterial community at 60 m was overall slightly more diverse than the surface community throughout the entire year.

The bacterial richness estimates (Chao1) also followed a seasonal pattern with extremes represented by late autumn/early winter versus bloom periods (Figure 6.3B). Mean bacterial richness estimates for the spring bloom were approximately 600 predicted OTUs. In the three surface sampling depths, richness clearly peaked in November and December with over 2000 predicted OTUs, whereas richness at 60 m remained relatively constant at 2000 estimated OTUs from May to November followed by a slight elevation in December. Overall, OTU richness estimations were 1000 species lower from May to October for 1, 5 and 10 m compared to 60 m.

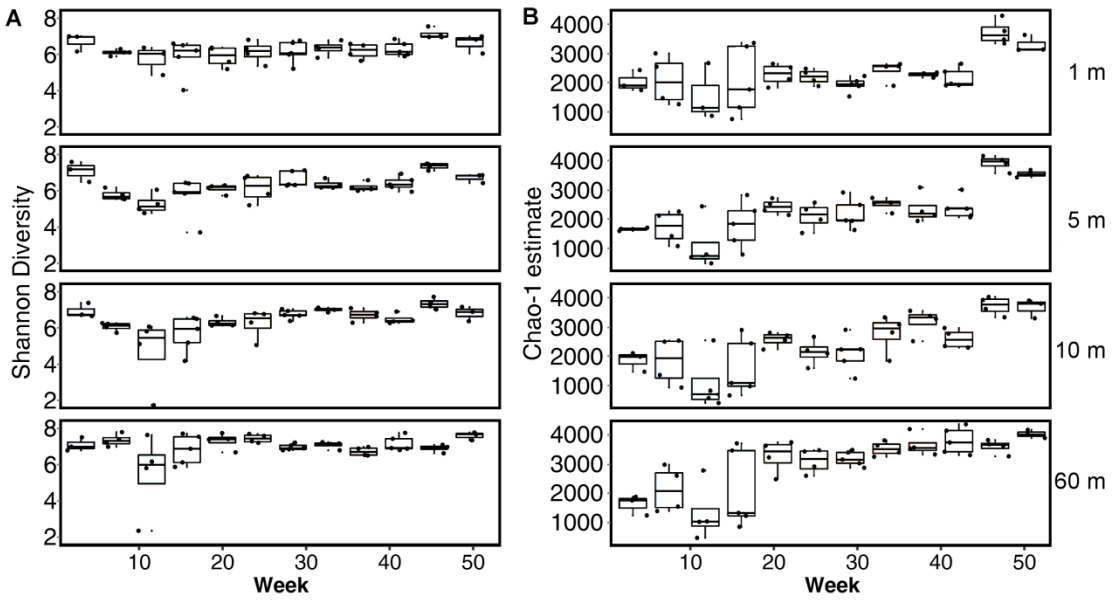


Figure 6.3: Alpha-diversity of 16S community in the Bedford Basin averaged monthly at each depth.

(A) Shannon diversity and (B) Chao1 estimates were calculated in QIIME. Weekly samples are plotted as dots over the year and monthly averages are represented as box plots. The ends of the box represent the 25th and 75th percentiles; the whiskers represent minimum and maximum range disregarding outliers.

6.3.4. Microbial community structure in space and time

We used Bray-Curtis dissimilarity to investigate the temporal variability in microbial community composition in the Bedford Basin (Figure 6.4 and Figure 6.5). Within the same depth, the average Bray-Curtis similarities were calculated for all weekly time-lags (Figure 6.4A). Within seasons, community composition was found to be most similar in surface samples. These are samples within 1-10 and 40-50 weeks of time lag. Highest similarity was found in adjacent weeks, which were on average 54 % similar to each other (Figure 6.5). Opposing seasons (26 week time lag) were the most dissimilar (30 % similarity). The

seasonal pattern was not pronounced in the 60 m samples, which displays highest similarity between samples of all four depths. Figure 6.4B compares Bray-Curtis similarity between consecutive weeks, which highlights rapid community shifts, especially during the spring bloom (week 16) and the first oceanic intrusion (week 10/11).

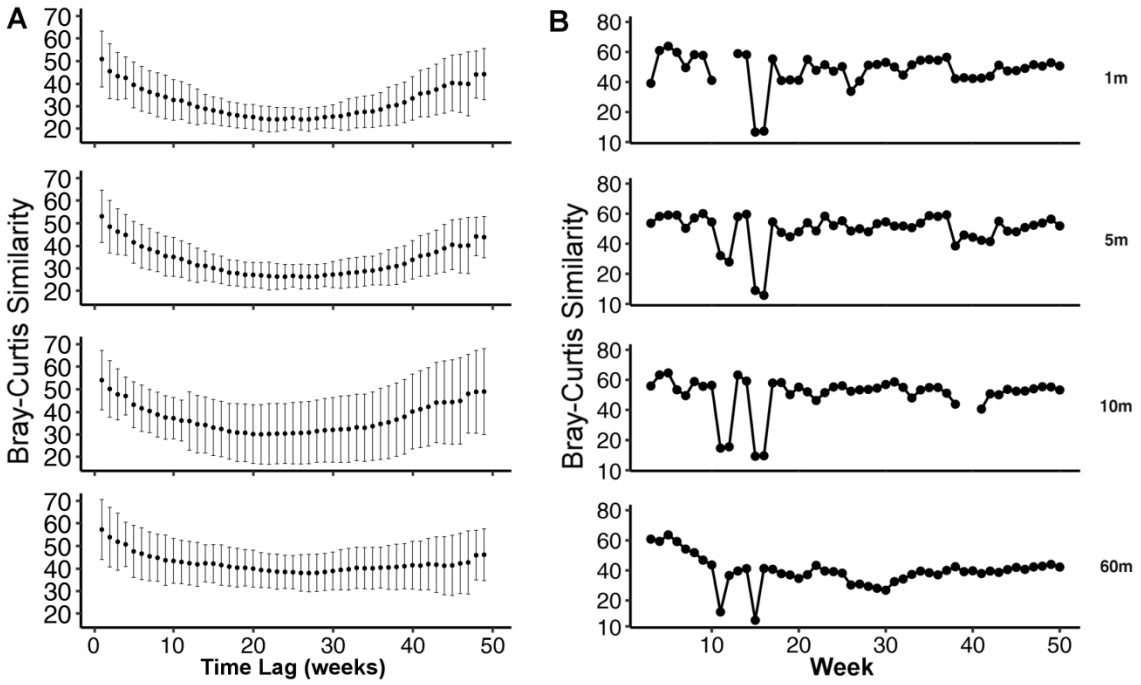


Figure 6.4: Seasonal pattern of Bray-Curtis similarity.
Panel A shows the similarity between samples for every depth based on Bray-Curtis similarity. Averages of every sample's Bray-Curtis similarity to every other sample at the same depth is displayed as a time lag series. Error bars show the standard deviation of each interval comparison. Panel B shows the one-weekly similarity changes over the year.

NMDS-analysis based on Bray-Curtis similarity visualizes the clustering of microbial communities for 1, 5 and 10 m depth (Figure 6.5). The most similar communities were from the same and adjacent months. Samples from January

until April were separated along the first axis, while the rest of the samples were additionally separated along the second axis. Over the weeks of summer and early autumn, clustering was less discrete possibly because community composition was governed by factors that were not captured in the 2-dimentional space. Extreme outliers occurred at 3, 12 and 16 weeks and were removed from Figure 6.5. Temperature and nutrient concentrations correlated most strongly with the two axes of community dissimilarity (Figure 6.5); nutrient concentration correlated with the first axis and temperature correlated with community similarity along the second axis. Overall, the samples at 60 m were much more similar to each other than in the surface; however, a slightly less distinct yearly community similarity cycle was visible (Figure 6.4). Outliers were from the spring bloom period in March.

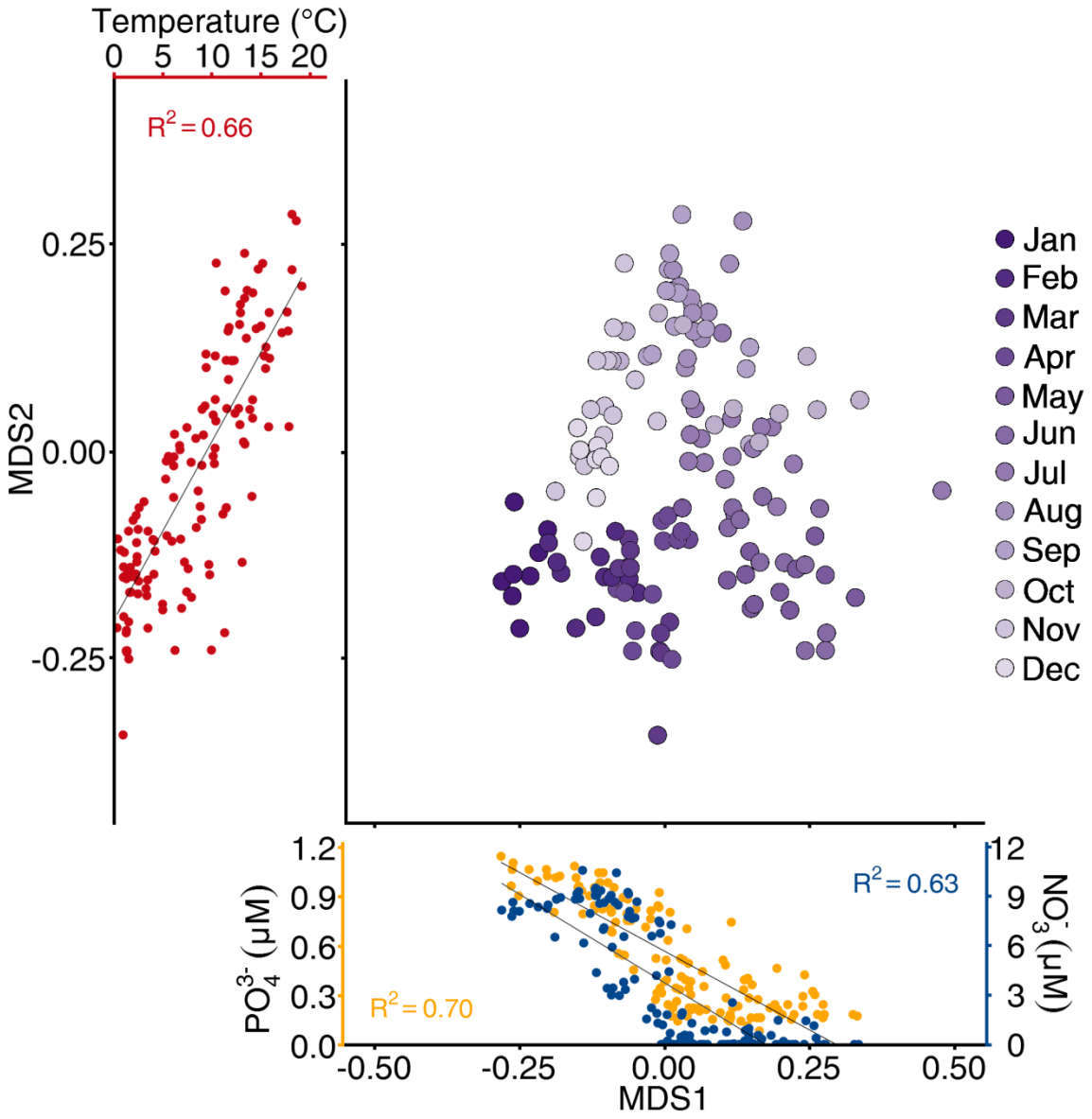


Figure 6.5: Community structure of weekly samples at 1, 5 and 10 m depth in the Bedford Basin in 2014.

Non-linear Multi-Dimensional Scaling (NMDS) was used to plot sample similarity according to their taxonomic composition and abundance. OTU counts were Hellinger transformed and NMDS plots were created based on Bray-Curtis dissimilarity between samples using *vegan* in R (R Core Team 2015; Dixon 2003). Each point represents one sample and is colour coded according to month. The distribution of samples was best correlated with PO_4^{3-} and NO_3^- along MDS1 and temperature along MDS2 (R^2=0.70, 0.63 and 0.66 respectively).

Community similarity shifts were also evident when directly observing the changes of the 100 most common OTUs over time (Supplemental Figure 19). The communities at 1, 5 and 10 m were very similar and followed similar patterns, whereas the community at 60 m showed a different pattern. The most common taxa represented in the Bedford Basin were from the phylum of Proteobacteria, followed by Bacteroidetes and Verrucomicrobia. At all depths, the disruptions at 12 (intrusion event) and 16 weeks (spring bloom) stand out. The second intrusion event at week 28/29 and autumn bloom period around week 38 were less distinct than the former two disruptions, and affects the 60 m community more profoundly than that found in the surface. The OTU pattern in the surface (1, 5 and 10 m) remained consistent during the first 18 weeks of 2014, except for the intrusion and spring bloom event. The following spring and summer weeks were much more variable, with some OTUs spiking in a single week. After the autumn bloom, the community returned to a more stable state. At 60 m, the community was much more constant over the seasons. Aside from the two disturbances in weeks 11 and 16, OTUs vary less substantially from one week to the next. Additionally, OTU shifts are prominent in the first 33 weeks and then stayed constant.

6.3.5. Patterns of OTU distribution

Over the year, OTUs displayed specific patterns of abundances with depth and time (Figure 6.6). Statistical analyses were carried out to determine what environmental factors most affected the occurrences of the 24 most common

OTUs, which were defined as accounting for more than 1% of reads at one depth (Supplemental Table 13). The factors considered were temperature, salinity, nitrate, phosphate, silicate, ammonium, chlorophyll, fluorescence, density, O_2, O_2 solubility and depth (Table 6.1, Table 6.2, Table 6.3 and Table 6.4). The most common OTUs in the surface (1, 5 and 10 m) were *Colwellia*, FlavobacteriaceaA, B and C, *Octadecabacter*, PelagibacteraceaeA and B, *Phaeobacter*, *PolaribacterA* and *B*, *Pseudoalteromonas*, *Psychrobacter*, RhodobacteraceaeA, B, C, D, E and F and *UlvibacterA* and *B*. At depth (60 m), the OTUs of *HTCC2207*, *SUP05*, RhodobacteraceaeG and *ZA3409c* also occurred in a significant fraction, while *Colwellia*, FlavobacteriaceaA and B, PelagibacteraceaeB, *Phaeobacter*, *PolaribacterA*, *UlvibacterA*, *Psychrobacter*, RhodobacteraceaeC, D, E and F made up less than 1% of reads (Figure 6.6).

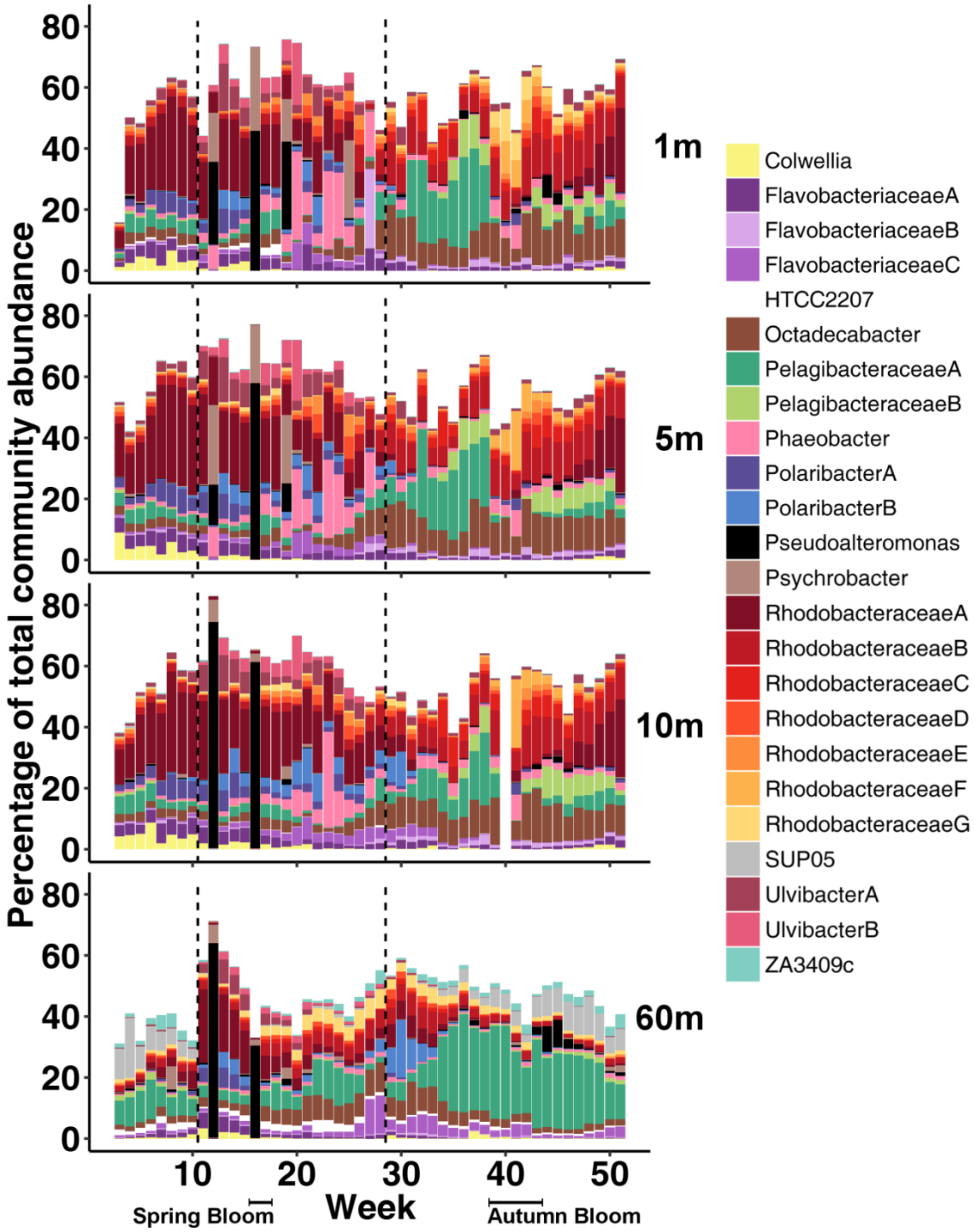


Figure 6.6: Most Abundant OTUs in the Bedford Basin in 2014.

Relative abundances of OTUs that made up more than 1% of all reads in at least one depth over the entire year are depicted for each depth and week. Rare taxa made up the missing portion to 100% of relative abundance. Intrusion events are indicated by dotted lines and peak bloom periods by horizontal bars on the x-axis.

The occurrence of each of the 24 most abundant OTUs was compared with environmental parameters and seasons using the BEST test, which showed that temperature (strongly correlated with O_2 solubility), depth, nutrient concentrations (nitrate, phosphate and silicate), chlorophyll, and seasons were significantly correlated with OTU distribution. However, parameters correlating with OTUs varied slightly in each depth (Table 6.1); the 1 m samples showed correlation with temperature, nutrients, fluorescence, density and season; 5, 10 and 60 m samples with O_2 concentration and salinity; 10 and 60 m with chlorophyll; and 10 m with ammonium concentrations (Table 6.1).

Table 6.1: Results of the BEST analysis for the most abundant OTUs.

	All depths	**1 m**	**5 m**	**10 m**	**60 m**
Significance[1)]	1%	1%	1%	1%	1%
Rho[1)]	0.558	0.437	0.551	0.606	0.680
Variables[2)]	Temperature[3)] Depth Nutrients[4)] Chlorophyll Season	Temperature Nutrients Fluorescence Density Season	Temperature Salinity O_2 concentration Season	Temperature Salinity Nutrients Ammonium O_2 concentration Chlorophyll Season	Temperature Salinity Nutrients O_2 concentration Chlorophyll Season

1) A Rho value with a significance level lower than 5% indicates that variables correlate significantly with OTU abundances.
2) Variables that correlate best with variations in OTU abundances are listed.
3) Temperature correlates with O_2 solubility, which is a derived variable of temperature and O_2 concentration
4) The group defined as nutrients contains highly correlated nitrate, phosphate and silicate concentrations

Since communities at 60 m were significantly different from surface communities, the OTU table was divided into surface and deep submatrices which were analyzed separately (Table 6.2). The LINKTREE analysis was used to identify thresholds that grouped each environmental variable into high and low values (Table 6.2). These groupings were then used to complete ANOSIM tests on the Bray-Curtis similarity matrix of the OTU table, which showed that almost all variables retrieved from the BEST analysis were correlated with OTU distribution at different degrees of significance (Table 6.2). In the surface, temperature, nitrate, phosphate, O_2 solubility and the seasons (except for winter) were most clearly associated with OTU distribution. At depth, silicate and O_2 concentrations were also significant. Chlorophyll, fluorescence, density and seasons were significant, but less clearly distinguishable, whereas salinity, ammonium and silicate in the surface, or fluorescence, density and O_2 solubility at depth were not significant (Table 6.2).

Table 6.2: Statistical comparison (ANOSIM) of the distribution of the 100 most common OTUs with environmental variables.

Environmental Variables	Thresholds[1]				R statistic		Significance level (%)[2]	
	1 m	5 m	10 m	60 m	Sur.[3]	Deep[4]	Sur.	Deep
Temperature (°C)	14.1	4.97	3.34	3.86	0.372	0.465	**0.1**	**0.1**
Nitrate (µM)	1.14	-[5]	9.18	9.03	0.180	0.355	**0.1**	**0.1**
Phosphate (µM)	0.33	-	0.37	1.33	0.153	0.689	**0.1**	**0.1**
Silicate (µM)	9.39	-	4.65	12.07	0.068	0.671	5.6	**0.1**
Chlorophyll	-	-	2.65	0.14	0.083	0.085	**4.4**	**1.5**
Fluorescence (mg m^{-3})	4.44	-	-	-	0.290	-	**0.2**	-
Density (kg m^{-3})	5.86	-	-	-	0.168	-	**1.7**	-
O_2 (µM)	-	312	339	274	0.153	0.62	**0.5**	**0.1**
Salinity (PSU)	-		30.45	30.9	0.081	0.07	14.2	16.5
O_2 solubility (mg L^{-1})	5.86	6.32	6.66	-	0.440	-	**0.1**	-
Ammonium (µM)	-	-	0.65	-	0.192	-	5.0	-
Depth		1 – 10 m vs 60 m			0.210		**0.1**	
Spring		Weeks 3 - 12			0.407	0.259	**0.1**	**0.7**
Summer		Weeks 13 – 25			0.238	0.154	**0.1**	**3.6**
Autumn		Weeks 26 – 38			0.294	0.257	**0.1**	**0.6**
Winter		Weeks 39 – 51			0.103	0.156	**0.8**	**4.9**

1) Thresholds were determined using the LINKTREE analysis.
2) An R value with a significance level lower than 5%, shown in bold, indicates that groupings are significantly different from each other. R increases with significance.
3) Sur. = Surface (1, 5 and 10 m)
4) 60 m
5) Blank values indicate that this variable was not significantly linked to OTU distribution according to the BEST analysis.

Finally, the association of the most abundant OTUs with parameters divided into high and low categories and seasons was determined with the SIMPER test (Table 6.3 and Table 6.4). *Pseudoalteromonas* and *Psychrobacter*, OTUs that occurred only during bloom periods (weeks 16 – 17 and 38 – 43), showed an overall dissimilarity/standard deviation smaller than one and are therefore not included in Table 6.3 and Table 6.4. OTUs that did not significantly contribute to the separation between environmental variables were also not included in Table 6.3 and Table 6.4. Generally, each OTU was found to dominate in one or two of

the seasons (Table 6.3 and Table 6.4, Figure 6.6). The surface community during the beginning winter months of 2014 was significantly populated by *Colwellia*, FlavobacteriaceaeA, *PolaribacterA* and RhodobacteraceaeA; RhodobacteraceaeA continued to be present during spring as well. The rest of the community shifted after the spring bloom and consisted of FlavobacteriaceaeB, *Phaeobacter*, *Polaribacter*B and *UlvibacterA* and B. Over the summer months, PelagibacteraceaeA, *Octadecabacter*, RhodobacteraceaeB and C became dominant in the community, the last three continuing to be significantly present in autumn. The community was disrupted at the beginning of autumn by the autumn bloom, after which PelagibacteraceaeB increased in abundance. Later in autumn an increase of RhodobacteraceaeA could be seen, returning the community to a state similar to that at the beginning of 2014 in late 2014.

Since the environmental conditions in the Bedford Basin in 2014 changed with the seasons, the most abundant OTUs were associated with the conditions present during those seasons (Table 6.3). In the surface, measurements during the winter and following the autumn bloom (week 43) showed that surface waters were overall cold (0.3 – 13.4°C, mean = 5.7°C), nutrient rich (nitrate = 1.3 – 10.6 µM, mean = 7.0 µM), low chlorophyll/fluorescence (0.46 – 23.95 mg/m^3, mean = 3.51 mg/m^3 and 1.14 – 18.13 mg/m^3, mean = 3.87 mg/m^3) and highly oxygenated (207 – 739 µM, mean = 307 µM). During the spring, nutrients were depleted in the surface (nitrate = 0 – 10.5 µM, mean = 1.1 µM) and water temperature increased due to solar heating and stratification (1.23 – 19.2°C, mean = 9.3°C).

Increased primary productivity was reflected by increase in chlorophyll and fluorescence (0.1 – 22.8 mg/m^3, mean = 4.94 mg/m^3 and 1.96 – 18.87 mg/m^3, mean = 5.89 18.87 mg/m^3), and O_2 concentrations were overall higher in the surface throughout spring and summer (214 – 403 µM, mean = 311 µM; Figure 6.2).

Table 6.3: Hellinger transformed abundances are listed for discriminatory OTUs in the surface (1, 5 and 10 m) that contributed to the overall dissimilarity between sample grouping pairs (dissimilarity/standard deviation >1) determined using SIMPER. Group pairs are defined using the thresholds in Table 6.2. The three OTUs that contributed most to the overall differences are in bold. Results for phylotypes with a dissimilarity/standard deviation below 1 are not displayed.

		Colwellia	**FlavobacteriaceaeA**	**FlavobacteriaceaeB**	***Octadecabacter***	**PelagibacteraceaeA**	**PelagibacteraceaeB**	***Phaeobacter***	***PolaribacterA***	***PolaribacterB***	**RhodobacteraceaeA**	**RhodobacteraceaeB**	**RhodobacteraceaeC**	***UlvibacterA***	***UlvibacterB***
Temperature	high	0.25	0.95	0.86	**2.4**	1.69	1.04	0.86	0.7	0.64	**1.52**	**2.25**	1.19	1.15	0.42
	low	1.21	1.56	0.77	**1.3**	1.42	0.27	0.65	1.54	0.89	**3.01**	**1.17**	0.05	1.42	0.95
Nitrate	high	1.05	1.37	0.51	1.78	**1.45**	0.86	1.00	1.25	0.46	**2.56**	**1.71**	0.62	1.3	0.43
	low	0.22	1.05	0.94	2.17	**1.67**	0.66	1.78	0.75	0.78	**1.51**	**2**	0.93	1.17	0.7
Phosphate	high	0.87	1.23	0.65	1.95	**1.61**	**0.87**	1.17	1.07	0.56	**2.21**	1.84	0.79	1.21	0.46
	low	0.17	1.12	0.9	2.09	**1.52**	**0.56**	1.84	0.82	0.78	**1.61**	1.92	0.81	1.25	0.77
Chlorophyll	high	-	0.93	1.04	1.85	**1.4**	**0.53**	1.55	0.85	1.01	**1.77**	1.66	0.75	1.23	0.74
	low	-	1.29	0.81	2	**1.68**	**0.98**	1.34	1.14	0.64	**2.45**	1.88	0.77	1.16	0.54
Fluorescence	high	-	0.95	-	**2.44**	**1.91**	0.93	1.22	0.47	0.2	**0.64**	2.31	1.4	0.92	0.14
	low	-	1.38	-	**1.88**	**1.42**	0.59	1.58	1.14	0.66	**2.35**	1.78	0.56	1.42	0.74
Density	high	-	1.25	0.73	1.91	**1.46**	0.73	1.46	1.06	-	**2.19**	**1.81**	0.61	1.3	0.69
	low	-	1.19	0.25	2.55	**1.97**	0.62	1.45	0.46	-	**0.53**	**2.43**	1.55	1.12	0.09
O_2	high	0.68	1.33	1.09	1.5	**1.3**	-	1.72	1.32	1.1	**2.84**	**1.33**	0.1	1.49	0.18
	low	0.56	1.11	0.74	2.19	**1.7**	-	1.35	0.88	0.59	**1.77**	**2.06**	1.03	1.16	0.77
O_2 Solubility	high	0.73	1.34	0.91	1.75	**1.37**	0.43	1.46	1.24	0.91	**2.54**	1.61	**0.38**	1.41	0.8
	low	0.16	0.86	0.4	2.68	**2.09**	1.47	1.43	0.35	0.06	**0.62**	2.56	**1.78**	0.87	0.8
Seasons	spring	0.48	1.19	1.24	1.25	0.94	-	1.91	1.27	**1.54**	**2.76**	**1.07**	0.01	1.68	**1.57**
	summer	0.15	1.03	0.75	2.56	**2.53**	0.87	1.4	0.45	0.46	**0.85**	2.35	**1.27**	0.96	0.24
	autumn	0.27	0.85	0.34	2.79	1.13	**1.66**	1.53	0.6	0.02	**1.25**	2.78	**1.79**	1.06	0.01
	winter	**1.61**	1.81	0.55	1.47	1.67	0.42	0.82	**1.78**	0.37	**3.26**	1.38	0.09	1.26	0.33

The community at depth (60 m) was significantly different from the surface and seasonality was less pronounced (Table 6.2). Until the spring bloom (16 weeks), *HTCC2207*, PelagibacteraceaeA, *SUP05* and *ZA3409c* dominated the bacterial community (Table 6.4). The spring bloom, which was followed by a community of *HTCC2207*, *Octadecabacter*, PelagibacteraceaeA, RhodobacteraceaeA as well as *UlvibacterB* disrupted the original community. The increase of relative abundance in *Phaeobacter* that was seen in the surface after the spring bloom was completely missing at depth (Figure 6.6). The occurrence of *Octadecabacter*, PelagibacteraceaeA, RhodobacteraceaeA and G continued over summer. The increase of FlavobacteriaceaeB and *PolaribacterB* in summer was more distinct at depth compared to the surface. *PelgibacteraceaeA* reached some of its highest relative abundances at the end of summer (Table 6.4, Figure 6.6). After the autumn bloom, the community was again dominated by PelagibacteraceaeA, *SUP05* and *ZA3409c*, moving, at the end of the year, towards very similar relative abundances to those seen at the beginning (Figure 6.6).

Table 6.4: Hellinger transformed abundances are listed for discriminatory OTUs in the deep (60 m) that contributed to the overall dissimilarity between sample grouping pairs (dissimilarity/standard deviation >1) determined using SIMPER. Group pairs are defined using the thresholds in Table 6.2. The three OTUs that contributed most to the overall differences are in bold. Results for phylotypes with a dissimilarity/standard deviation below 1 are not displayed.

		FlavobacteriaceaeC	***HTCC2207***	***Octadecabacter***	***PelagibacteraceaeA***	***PolaribacterB***	***RhodobacteraceaeA***	***RhodobacteraceaeB***	***RhodobacteraceaeG***	***SUP05***	***UlvibacterB***	***ZA3409c***
Temperature	high	1.31	0.93	1.67	**3.15**	0.57	**0.93**	1.49	1.27	**2.02**	0.08	1.29
	low	1.08	1.14	1.88	**1.9**	0.98	**20.3**	1.64	1.47	**0.47**	0.89	0.7
Nitrate	high	1.12	0.98	1.49	-	**0.3**	**0.95**	1.33	1.12	**2.27**	0.12	1.39
	low	1.3	1.04	1.99	-	**1.11**	**1.75**	1.74	1.55	**0.64**	0.66	0.76
Phosphate	high	1.27	1.1	1.83	**2.94**	0.53	**1.16**	1.62	1.44	**1.69**	0.37	1.22
	low	0.97	0.67	1.46	**1.37**	1.61	**2.31**	1.3	0.97	**0.11**	0.62	0.33
Silicate	high	1.25	0.9	-	**2.96**	0.48	**1.14**	1.6	1.39	**1.72**	0.56	1.22
	low	1.09	1.12	-	**1.47**	1.67	**2.27**	1.4	1.31	**0.14**	0.28	0.42
Chlorophyll	high	1.27	0.65	-	**2.38**	1	**1.56**	-	0.62	**1.07**	0.86	0.89
	low	1.16	1.05	-	**2.87**	0.5	**1.22**	-	1.42	**1.66**	0.37	1.2
O_2	high	0.63	0.96	1.09	**1.27**	1.19	**2.55**	0.97	1.36	**0.14**	0.41	0.2
	low	1.27	1.17	1.82	**2.76**	0.7	**1.27**	1.61	1.33	**1.5**	0.42	1.13
Seasons	spring	0.88	1.31	1.91	**2.06**	0.96	**2.07**	-	1.5	**0.45**	1.26	0.54
	summer	1.89	0.87	2.26	**2.61**	**1.4**	1.53	1.98	1.77	**0.81**	0.34	1.12
	autumn	1.23	0.7	1.4	**3.57**	0.29	**0.27**	1.24	1.11	**2.42**	0.05	1.31
	winter	0.72	0.91	-	-	**0.12**	**1.59**	-	0.91	**2.08**	0.12	1.29

To examine the relationship between the bacterial community at the surface (1, 5 and 10 m) with the community at depth (60 m), which can be linked through mechanisms such as sinking particles, faecal pallets or by resuspension of the sediments, a time lag analysis was performed (Figure 6.7). OTUs were correlated based on their abundances in the surface compared to depth. The time lags showed that most of the abundant OTUs occurred first in higher relative abundances in the surface and later at lower abundances at depth (positive time lag and positive correlation), which is indicative of sinking. OTUs that occurred first in the surface included *Colwellia*, FlavobacteriaceaeA and C, *HTCC2207*, PelagibacteraceaeA and B, *Phaeobacter*, *PolaribacterA* and *B*, most of the Rhodobacteraceae, and *UlvibacterA* and *B*. Of those, only FlavobacteriaceaeC, PelagibacteraceaeA and *B*, *PolaribacterB* and RhodobacteraceaeC and *G* reach higher abundances at depth compared to the surface. OTUs that displayed higher abundances at depth were FlavobacteriaceaeB, *Octadecabacter*, RhodobacteraceaeB, *SUP05* and *ZA3409c*. All these OTUs increased in the surface following higher abundances at depth.

High positive values without time lag (0 weeks) point to OTUs that are abundant throughout the entire water column at the same time, which can be seen for the specific OTUs that spike during the spring bloom (*Pseudoalteromonas*, *Psychrobacter* and *UlvibacterA*) and that generally occurred over the weeks of the spring bloom (FlavobacteriaceaeA, *PolaribacterA* and *B*, *UlvibacterB*).

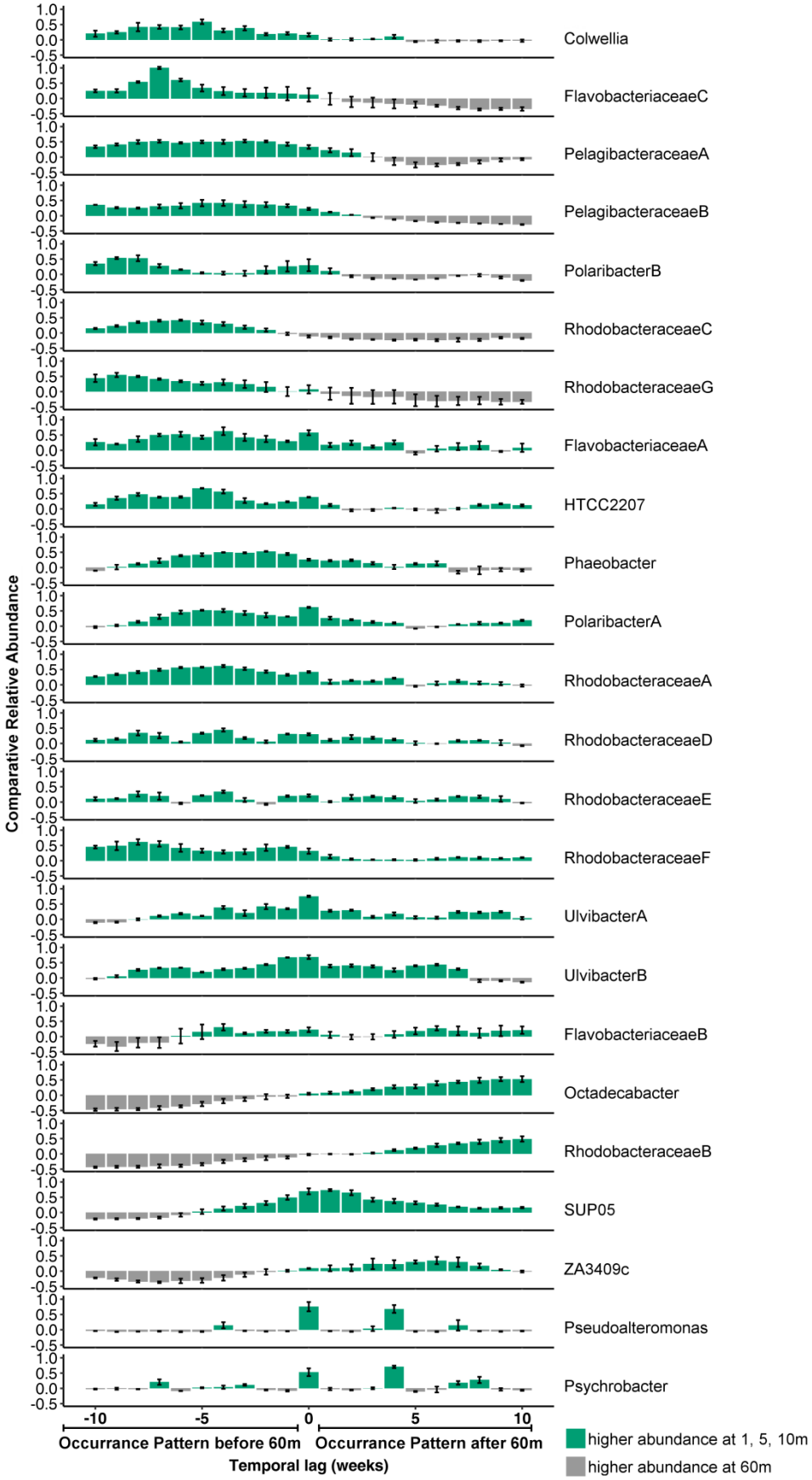

Comparative Relative Abundance
1.0
0.5
0.0
-0.5
Colwellia
FlavobacteriaceaeC
PelagibacteraceaeA
PelagibacteraceaeB
PolaribacterB
RhodobacteraceaeC
RhodobacteraceaeG
FlavobacteriaceaeA
HTCC2207
Phaeobacter
PolaribacterA
RhodobacteraceaeA
RhodobacteraceaeD
RhodobacteraceaeE
RhodobacteraceaeF
UlvibacterA
UlvibacterB
FlavobacteriaceaeB
Octadecabacter
RhodobacteraceaeB
SUP05
ZA3409c
Pseudoalteromonas
Psychrobacter
-10
-5
0
5
10
Occurrance Pattern before 60m
Occurrance Pattern after 60m
Temporal lag (weeks)
higher abundance at 1, 5, 10m
higher abundance at 60m

Figure 6.7: Relationship of taxa between surface (1, 5 and 10 m) and deep (60 m) samples

The occurrence of taxa in the surface samples compared to the 60 m sample was analyzed with the Applied Statistical Time Series Analysis (*astsa*) package in R (Stoffer 2012; R Core Team 2015). Taxa occurred either earlier or later than in the 60 m samples (-10 to +10 weeks were considered). Positive or negative correlations are indicated by colour (positive = green, negative = grey), and the strength of correlation is marked by the y-axis. The positive time lag looked at OTUs occurring in the surface after they occurred at depth; with a positive correlation indicating a higher relative abundance in the surface after a lower relative abundance at depth and a negative correlation indicating a lower relative abundance in the surface after a higher relative abundance at depth. The negative temporal lag looked at the OTUs that occurred in the surface before they occurred at depth. A positive correlation indicated higher abundances in the surface before a lower abundance at depth and a negative correlation indicated a lower abundance in the surface before higher abundance at depth.

6.3.6. Blooms and Intrusions

The blooms in 2014 occurred around weeks 16 and 40 in the Bedford Basin. The spring bloom changed the bacterial community dramatically (Figure 6.6 and Supplemental Figure 19). In week 16, all four depths saw a shift away from a community dominated by FlavobacteriaceaeA and B, *Octadecabacter*, PelagibacteraceaeA, *PolaribacterB*, RhodobacteraceaA, and *UlvibacterA* and *B* to a community where *Pseudoalteromonas* and *Psychrobacter* comprise 73, 77, 61 and 32% of all reads for 1, 5, 10, and 60 m, respectively. In week 17, the community of OTUs making up more than 1% of reads returns to a very similar composition to that of week 15, although an increase in relative abundance of *Phaeobacter* and PelagibacteraceaeA was seen (Figure 6.6).

The autumn bloom, which peaked in week 38, did not display the drastic shifts in OTUs seen during the spring bloom. Rather, the relative abundance of PelagibacteraceaeA and B drastically decreased. PelagibacteraceaeA decreased from 18.8 to 0.0%, 22.8 to 0.1% and 23.8 to 0.2% in the three surface depths (1, 5 and 10 m respectively), and PelagibacteraceaeB decreased from 7.0 to 0.0%, 7.8 to 0.1% and 8.6 to 0.2% from week 38 to week 41 (Figure 6.6). OTUs that increased in relative abundance in the surface over the autumn bloom period were *Phaeobacter*, *PolaribacterA*, *Pseudoalteromonas* and RhodobacteraceaeE and F. The autumn bloom did not cause disruptions in the most abundant OTUs 60 m, except for a decrease in PelagibacteraceaeA (25.7 to 20.6%) and an increase in *Pseudoalteramonas* that was seen during the late bloom stage.

As evidenced in salinity and temperature profiles (Figure 6.2), there were two intrusions of oceanic waters, the first from week 10 to 11, and the second from week 28 to 29. The earlier intrusion caused mixing of the entire water column, leading to a more even community throughout all depths in week 11, followed by a bloom of *Phaeobacter*, *Pseudoalteromonas* and *Psychrobacter* in week 12 (Figure 6.6). The second intrusion did not alter the community as radically as the first one (Figure 6.6). At 1 and 5 m, *Octadecabacter* and RhodobacteraceaeA increased, while at 10 m and 60 m *PolaribacterB* increased. At 60 m a decrease of Flavobacteriaceae was seen (Figure 6.6).

6.3.7. Diazotrophs in the Bedford Basin

Along with the total bacterial community, the diazotrophic community was also investigated on four dates (March 19th, June 18th, September 24th and December 17th 2014). Samples that tested positive for *nifH* were March, 5 and 60 m; June, 1, 10 and 60 m; September, 10 and 60 m; and December,1, 10 and 60 m. After quality control, 162,148 reads were recovered, which included 1122 *nifH* OTUs. The diazotrophic community in the Bedford Basin changed with depth and time (Figure 6.8 and Figure 6.9). All samples were dominated by sub-cluster Ip sequences, sequences that are dominated by proteobacteria (Zehr et al. 2003; Figure 6.8 and Figure 6.9). Cyanobacterial sequences were only found in the surface (1 and 10 m) in June and September and, with the exception of one OTU in September (92% identity to *Calothrix)*, all were assigned to the *Candidatus* Atelocyanobacterium thalassa clade 2 (Figure 6.9). Clusters II and III were found in all samples except for the 5 m sample from March (Figure 6.8).

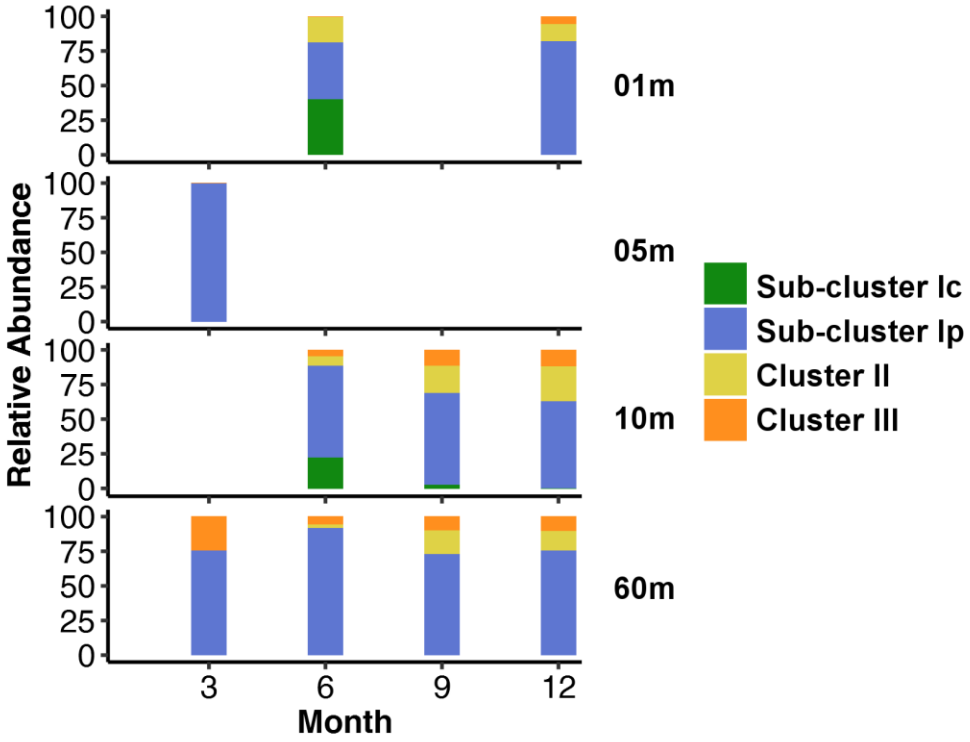


Figure 6.8: Distribution and relative abundances of *nifH* clusters throughout the water column of the Bedford Basin in 2014.

Relative abundances of each *nifH* cluster are shown for samples collected in the Bedford Basin according to depth (1, 5, 10 and 60 m) in 2014.

Two of the three most abundant OTUs were newly clustered during the bioinformatic analyses. The three most abundant sequences made up 9.9, 8.7 and 8.6% of all *nifH* sequences and were all sub-cluster Ip sequences (Figure 6.9). The most abundant OTU (New.ReferenceOTU7), comprising 99% of all sequences in the 5 m sample from March, was 92% identical to a sequence previously isolated from the Tibetan Plateau (Zhang et al. 2006). New.ReferenceOTU5, which was 99% identical to a sequence previously identified in the Laptav Sea and below sea ice (Fernández-Méndez et al. 2016), occurred in all samples, with highest abundances at 60 m (44, 22, 14 and 3.7% in

March, June, September and December respectively). The third most abundant sequence was previously detected in marine sediments of Narragansett Bay (Fulweiler et al. 2013) and was here most prominent at 60 m.

The samples collected on December 16th at 1 and 10 m showed the overall greatest diversity. They contained 440 and 368 unique OTUs respectively (Figure 6.9), whereas the 5 m sample from March 19 showed one of the lowest diversity, with New.ReferenceOTU5 making up 99% of all reads (closest reference genome was *Desulfovibrio alkalitolerans*, 77% sequence similarity).

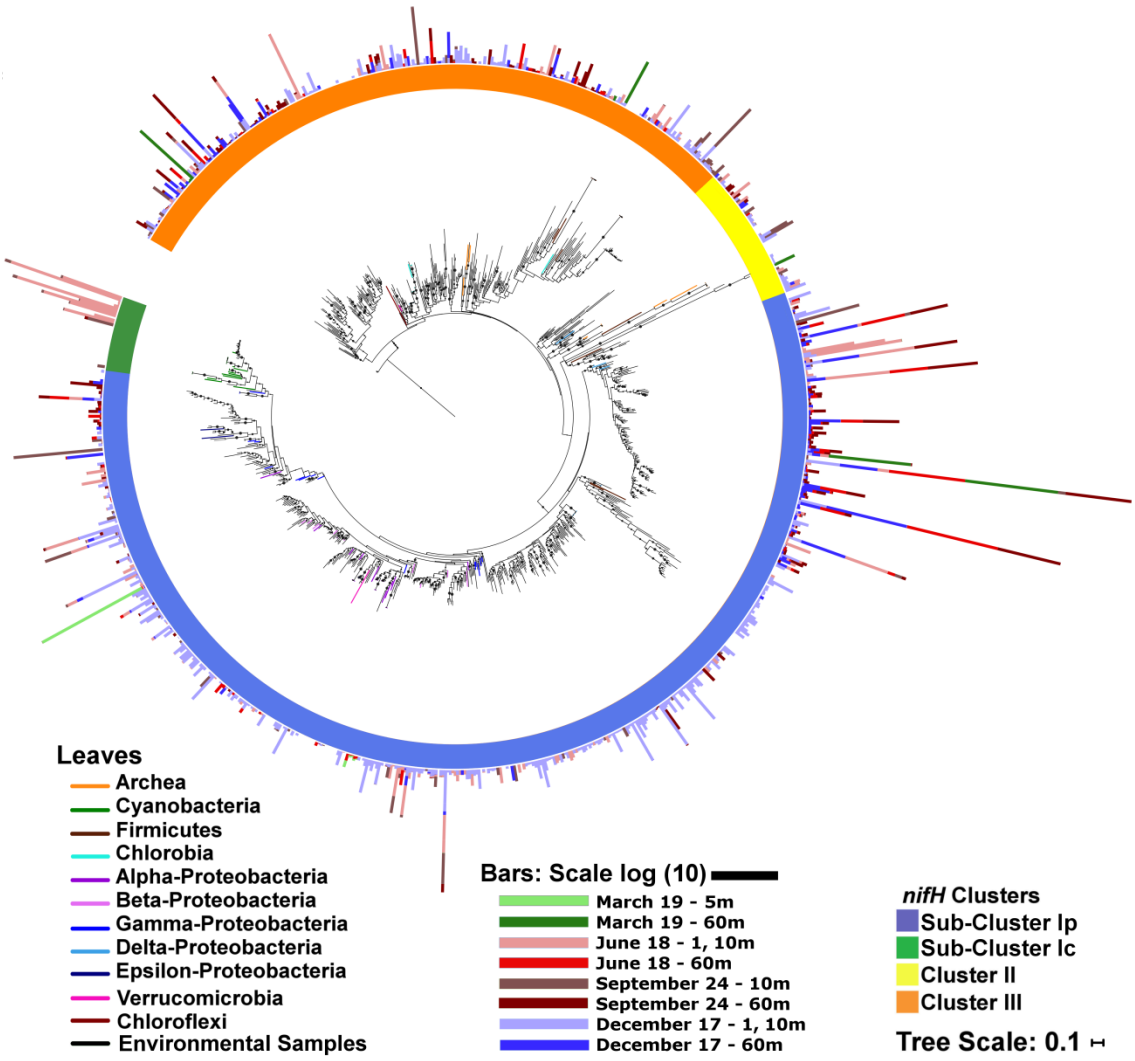


Figure 6.9: Phylogenetic *nifH* diversity and abundances in the Bedford Basin.

768 OTUs were recovered through high-throughput sequencing of the *nifH* gene in the Bedford Basin. The phylogenetic association of these OTUs and reference sequences was inferred by using the maximum likelihood method based on the GTR-GAMMA model after codon-aligning sequences (MAFFT v. 7; Yamada et al. 2016; Stamakatis 2014; Katoh et al. 2002). Bootstrap values were calculated from 100 tree replicates and values >50% are shown as dots. The tree was displayed with branch lengths showing the number of substitutions per site. Leaves of reference genomes are coloured according to their taxonomy. Black leaves indicate environmental sequences. The outer bars indicate the logarithmically transformed number of sequences present for each OTU in the OTU table rarefied to 1500 reads. Bars are coloured based on season (spring: green, summer: red, autumn: brown, winter: blue) and depth of their detection (1,

5 and 10 m: light, 60 m: dark). Clusters are assigned according to Zehr et al. (2003).

6.4. Discussion

Microbial communities in the ocean display great diversity and genetic potential across temporal and spatial scales (Sunagawa et al. 2015; De Vargas et al. 2015; Gilbert et al. 2012). As a whole, they contribute to approximately half of all biogeochemical processes on the planet and play a substantial role in the global climate (Arrigo 2005). As the ocean changes due to anthropogenic influences, these microbial communities have already started to respond (e.g. Bunse et al. 2016; Hutchins et al. 2009). Efforts to capture the current state of marine microbial communities and to monitor ongoing changes have been increasing concurrent with the decrease of high throughput sequencing costs (reviewed by Fuhrman et al. 2015). The Bedford Basin presents an ideal site for a highly resolved microbial time-series study. Although there has been continuous sampling of oceanographic parameters in the Bedford Basin since 1992, there is only a single five-year study of the microbial community with biweekly samples at 1 m conducted from preserved historical samples (El-Swais et al. 2014). Extending the weekly sampling effort to include regular microbial analysis provides an unparalleled opportunity to monitor the changing ocean. The Bedford Basin experiences low O_2 conditions at depth during late autumn and winter weeks, which may provide insights into community responses that occur as oxygen minimum zones (OMZ) expand in oceans affected by climate change (Stramma et al. 2008). In addition, short-term anthropogenic or natural

disruptions can be investigated at temporally- and spatially-resolved spans. This study will provide a baseline for future work in the Bedford Basin to which changes, both natural and anthropogenic, can be related.

6.4.1. Environmental influences on the microbial community

Overall, the patterns of microbial community distribution appeared well resolved by the weekly sampling regime (Figure 6.6). Marine microbial doubling times in the temperate oceans have been shown to range from 0 to a maximum of 2 per day during the spring bloom, indicating that weekly sampling should be sufficient to resolve patterns over the year (Huete-Stauffer et al. 2015). This is supported by our Bray-Curtis similarity analyses, which identified smooth transitions except in the cases of extreme disruptions as seen during the first oceanic intrusion event and weeks of the spring bloom (weeks 11, 16; Figure 6.4, Figure 6.6, Supplemental Figure 19). To investigate the transitions during these conditions more closely, a much more frequent sampling regime, such as one that is daily for a short period, might be more insightful. The alpha-diversity (the evenness measure of Shannon diversity and richness estimator of Chao1) in the Bedford Basin followed the same pattern recorded by El-Swais et al. (2014), which was similar to other temperate ocean time series (Giovannoni and Vergin 2012; Gilbert et al. 2011; Figure 6.3). Diversity was highest during mixed, nutrient rich winter conditions, lowest during the spring and autumn blooms, and generally unchanging throughout late spring and summer. There are multiple factors that may contribute to this pattern. High diversity during cold weeks could result from

mixing species that usually reside at depth up to the surface when the water column is not stable. Among the most abundant OTUs, this migration was seen for *ZA3409c*, which was almost exclusively found at depth during the stratified summer, but showed increased numbers in the surface in late autumn and winter (Figure 6.6); this was also reflected in the time lag analysis where greater abundance at depth preceded increasing abundance in the surface (Figure 6.7). Samples at 60 m, being overall more diverse, could hence contribute to a higher overall diversity in the surface during mixed winter conditions. High diversity in winter could also stem from high nutrient conditions supporting a more diverse community. Few taxa are adapted to succeed in the low nutrient regime of the stratified warm period, and thus thrive only in cold, nutrient-rich conditions. This is widely observed when comparing nutrient rich ocean water masses to the nutrient-limited surface ocean (Sunagawa et al. 2015; Ghiglione et al. 2012). The very low diversity seen during the spring bloom was attributed to the spiking OTUs (*Phaeobacter*, *Pseudoalteromonas* and *Psychrobacter*) which decreased the evenness of the samples by decreasing single reads, a factor used in the estimation of richness in the sample.

Similar microbial communities were identified using the Bray-Curtis similarity measure. The fact that surface samples (1, 5 and 10 m) of the same or adjacent weeks were most similar was expected (Figure 6.4 and Figure 6.5). Similar environmental conditions throughout the surface and the proximity of the three surface depths likely allowed species exchange between depths despite stratification (Cram et al. 2015a). However, the community transitioned from

week to week with changing environmental conditions. The variables that correlated most with community dissimilarity were temperature and nutrient concentration (Table 6.1 and Figure 6.5). Since temperature is the main driver for stratification and consequently nutrient depletion in the surface, it can be considered as the strongest determining factor for community composition in the surface of the Bedford Basin. Similar conclusions were drawn by El-Swais et al. (2014b), along with other temperate time-series and global ocean studies (Cram et al. 2015; Lucas et al. 2015; Sunagawa et al. 2015; Chow et al. 2014; Raes et al. 2011). Temperature as the major driver of community composition is also supported by the fact that the bacterial communities of the Bedford Basin showed high similarity at beginning and end of the year when temperatures were comparatively low (Figure 6.4, Figure 6.5 and Supplemental Figure 19). Since ocean temperatures are likely to increase with climate change, we can expect to see changing microbial communities as a result.

Apart from the seasonal large-scale changes, there were several other environmental factors associated with community composition. Similar to El-Swais et al. (2014), community structure was significantly linked to chlorophyll concentration and fluorescence, suggesting a possible relationship between primary production and the bacterial community (Table 6.1 and Table 6.2); explanations for this may include bacterial dependence on phytoplankton-derived material for energy and nutrient acquisition (Kujawinski et al. 2016, Fouilland et al. 2014; Aota et al. 2001), symbiosis (De Vargas et al. 2015), and/or dependence on similar environmental parameters (Steele et al. 2011).

Dysoxic levels that mark the initiation of anaerobic respiration were reached in late autumn and winter at 60 m (31 – 90 μM; Figure 6.2), coinciding with an increase in the relative abundance of *SUP05*, a species recorded to favor low O_2 conditions (Wright et al. 2012; Walsh et al. 2009). In contrast, throughout spring and summer, O_2 concentrations and the bacterial community were both much more homogeneous across the entire water column. Slight variations in O_2 concentrations in the surface correlated with chlorophyll, possibly linking the O_2 photosynthesis to phytoplankton that then links O_2 to surface bacterial communities through a secondary effect. O_2 concentration was correlated with community composition at 5, 10 and 60 m; while O_2 concentration is key to the switch from aerobic to anaerobic respiration and hence influences microbial communities by selecting for obligate aerobic or anaerobic organisms, it may also be an important factor for community structure at much higher concentrations, since O_2 levels initiating anaerobic respiration were never reached in the surface of the Bedford Basin (Spietz et al. 2015). In a global sampling regime, O_2 was shown to be an important factor for community similarity in surface samples, suggesting that bacteria respond to a great range of oxygen concentrations (Sunagawa et al. 2015).

6.4.2. *Seasonal patterns of the dominant microbial community*

One aim of this study was to establish a high resolution baseline of the most abundant OTUs that can be used for comparison with future microbial investigations in the Bedford Basin. Dominant OTUs were defined as comprising

at least 1% of reads in at least one depth, a standard which has been used previously to differentiate abundant and rare taxa (Lynch and Neufeld 2015; Forth et al. 2014). The microbial community in the Bedford Basin has been monitored by BIO through flow cytometry for over 20 years (Li 2014; Li and Harrison 2008; Li, Harrison and Head 2006; Li and Dickie 2001), and more recently a proteomics study and a bacterial diversity study were conducted in the Basin providing a good base for further investigations (El-Swais et al. 2014; Georges et al. 2014).

6.4.2.1. The Surface Community

At the beginning of 2014, the dominant winter community (*Colwellia*, FlavobacteriaceaeA, *PolaribacterA* and RhodobacteraceaeA) was composed of OTUs that have previously been associated with cold conditions. *Colwellia* was almost exclusively found in the winter's surface waters. Species of *Colwellia* have been isolated from the Arctic Ocean; growth has been recorded at temperatures down to -12°C, and the excretion of stabilizing proteins and active enzymes at low temperatures was observed (Marx et al. 2009; Methe et al. 2005; Huston et al. 2000, 2003, 2004). Interestingly, the genome of one species of *Colwellia* was sequenced during the Deepwater Horizon spill as an organism taking part in hydrocarbon degradation (Mason et al. 2015). Since the Bedford Basin is an active commercial harbour and serves as a catchment for road run off, it is contaminated with a variety of hydrocarbons that may provide a substrate for *Colwellia* (Hellou et al. 2002). On the other hand, *Polaribacter* species have been isolated from both Arctic and temperate oceans, and are widely distributed

throughout those regions (Nedashkovskaya et al. 2013; Yoon et al. 2006; Gonski et al. 1998). Two species of *Polaribacter* identified during the spring bloom period in the North Sea have often been found in association with algae and were shown to degrade polysaccharides (Xing et al. 2015). This agrees with our findings for *PolaribacterB*, which increases in relative abundance during the spring bloom and remains high in number throughout most of spring when diatoms are abundant (Li et al. 2008). One of the *Polaribacter* species isolated by Xing et al. (2015) displayed the genetic potential to express photorhodopsin, which may be the case for *PolaribacterA* in the Bedford Basin, since it is almost exclusively found in surface samples indicating a preference for the euphotic zone.

The other two abundant OTUs were of the families Flavobacteriaceae and Rhodobacteraceae. These families contain numerous genera, which does not allow for many conclusions about specific metabolic traits. There was a progressive change in various OTUs from both families throughout 2014 in the Bedford Basin. RhodobacteraceaeA was the most abundant Rhodobacteraceae in winter and spring, while RhodobacteraceaeB and C became more abundant in summer and early autumn, with some Rhodobacteraceae specific to the autumn bloom (Figure 6.6). This demonstrates that Rhodobacteraceae strains have adapted to very different environmental conditions (Fu et al. 2013). Nevertheless, both families have been frequently found in productive, nutrient-rich environments and in temperate ocean time-series studies (El-Swais et al. 2014; Klindworth et al. 2014; Williams et al. 2013; Gilbert et al. 2011). Marine *Flavobacteria*, to which

FlavobacteriaceaeA belong, have generally been described to degrade biopolymers, and are found in association with algae and present in phytoplankton blooms (Klindworth et al. 2014; Dong et al. 2012; Thomas et al. 2011).

Following the spring bloom, the microbial community became less diverse (Figure 6.3), possibly as a result of the increasingly warmer, stratified, and nutrient-depleted waters in the surface. OTUs that were significantly associated with these conditions included FlavobacteriaceaeC, *Phaeobacter*, *PolaribacterB*, *UlvibacterA* and *B*. As previously discussed, FlavobacteriaceaeA and *Polaribacter* have been associated with algae blooms, and the increased relative abundance of FlavobacteriaceaeC and *PolaribacterB* following peaks in chlorophyll at 11, 15/16 and 19 weeks support a possible connection between those OTUs and chlorophyll-containing organisms (Xing et al. 2015; Klindworth et al. 2014).

The genus *Phaeobacter* has commonly been found in the surface of a variety of temperate marine regions (Figure 6.6; Gram et al. 2010; Pommier et al. 2006). One species, *Phaeobacter gallaeciensis*, is capable of producing the antibiotic tropodithietic acid, and has thus been studied as a potential health-promoting species in fish and mollusc farms, (Porsby et al. 2008). The ability to produce antibacterial compounds seems to be common in various *Phaeobacter* species, giving rise to the proposal that some *Phaeobacter* live in symbiosis with marine eukaryotes, participating in nutrient exchange while keeping harmful bacteria at bay (Buchan et al. 2005; Joint et al. 2002). This association could be occurring in

the Bedford Basin, as *Phaeobacter* was most abundant during the spring time, including during the spring bloom, a time when diatoms are thriving (Li and Dickie 2001).

UlvibacterA and *B* were present throughout and after the spring bloom. *Ulvibacter* are pigmented and were isolated from temperate coastal waters in Japan and China (Baek et al. 2014; Nedashkovskaya et al. 2004). They have since been found in other temperate and Arctic environments, and form part of the late bloom community in the North Sea, which also seems to be the case in the Bedford Basin (El-Swais et al. 2014; Bowman et al. 2012; Teeling et al. 2012).

The summer conditions in the Bedford Basin were stratified, nutrient depleted and warm (10 – 19°C in the surface). PelagibacteraceaeA was significantly associated with these summer conditions. PelagibacteraceaeB, *Octadecabacter* and RhodobacteraceaeB and C were also dominant during this time, although their abundances also remained high through autumn (Figure 6.6).

PelagibacteraceaeA (97% identity to reference genome *Candidatus* Pelagibacter ubique HTCC1062, NC 007205) might be the most abundant organism on earth and has adapted to ultra-oligotrophic conditions through its very small size (Giovannoni et al. 2005; Morris et al. 2002). This adaptation might explain the dominance of Pelagibacter to thrive in the Bedford Basin when nutrient concentrations are drawn down during the summer months.

Octadecabacter is a psychrophilic genus that has been found throughout the Arctic, Antarctic and temperate oceans (Vollmers et al. 2013). Thus far, *Octadecabacter* has not been a major focus of research; however, we can record

its presence in the Bedford Basin at the surface throughout summer and autumn, and at depth through the spring and summer in relatively constant abundances.

Except for a drastic decrease in relative abundances of PelagibacteraceaeA, possibly because of the loss of advantage of well suited adaptations to low-nutrient regimes, other OTUs were constant before and after the autumn bloom until eventually returning to a very similar community profile to that seen at the start of 2014.

6.4.2.2. The Deep Community

When studying microbial communities at depth, the true deep community cannot be entirely disentangled from bacteria that have sunk from the surface layers or have been resuspended from bottom sediments. Dominant OTUs that were significantly associated with deep samples were assumed to have habitats at 60 m, whereas others may have been transported there or were present in both environments. OTUs significantly associated with 60 m were *HTCC2207*, RhodobacteraceaeG, *SUP05* and *ZA3409c*. Although these organisms displayed distribution patterns over the year, seasonality was not a clear distinguishing factor (Table 6.2, Supplemental Figure 18). Rather, the driving factor was likely the O_2 concentration that was slightly correlated with season, which might explain a secondary correlation (Table 6.4).

HTCC2207 has been isolated from marine coastal environments, which is in agreement with our study area; however, they have been shown to express

proteorhodopsin (Stingl et al. 2007; Cho and Giovanni 2004). This would indicate that *HTCC2207* may have a preference for the euphotic zone since it could use solar radiation for energy conservation. However, we found *HTCC2207* predominantly at depth, possibly indicating the presence of enough carbon sources to generate energy through oxidative phosphorylation. Abundances of *HTCC2207* were highest after the spring bloom when a great variety of carbon sources should be available from sinking material.

The of autotrophic *SUP05* was significantly associated with winter and autumn, when O_2 concentrations at depth were dysoxic. *SUP05* has been widely identified through the temperate oceans where it is linked to low O_2 conditions (Glaubnitz et al. 2009, 2010; Walsh et al. 2009). It has the potential for nitrogen, hydrogen and sulfur cycling (Anantharaman et al. 2013; Walsh et al. 2009; Sunamura et al. 2004), thus future calculations of nutrient cycling budgets in the Bedford Basin will need to account for O_2 -limited respiration.

There are no cultured representatives of *ZA3409c* to our knowledge, but it has been found in the marine environment, including the Mediterranean Sea and the Mexican Gulf during the Deep Water Horizon oil spill, which may suggest the potential for hydrocarbon degradation, which are present in the Bedford Basin (Mason 2014; Moeseneder et al. 2005).

As discussed previously, Rhodobacteraceae is a very diverse group of organisms and it is very likely that different OTUs from this group will have separate niches. As seen in the Bedford Basin, these niches are present within the same water system and varying depths and seasons (Figure 6.6). RhodobacteraceaeG was

significantly present at depth over spring and summer, possibly indicating a better adaptation to high O_2 concentrations.

Of particular note is the community shift that occurred at 60 m from week 10 to 11, and to a lesser extent from weeks 28 to 29. This is unlikely to be the natural progression of the community, but rather the result of complete or partial water column mixing possibly resulting from an oceanic water intrusion. There are several observations that indicate an Atlantic Ocean water intrusion into the Bedford Basin between weeks 10 and 11 as well as weeks 28 and 29. Community changes seem to be driven by changing environmental conditions, most prominently temperature and nutrient concentration in the surface, and O_2 concentration at depth.

6.4.2.3. Interactions of surface and deep communities

Time lag correlation (Figure 6.7) revealed interactions between the communities at depth and at the surface, which are thought to be propagated through sinking, mixing and stratification (intrusion and bloom events are discussed below). Since the community composition in the three surface depths was not significantly different, these results were combined. Figure 6.7 depicts the relationship between surface and depth. Distinguishing true residents at depth from sinking or mixed material may be important when considering metabolic activity at depth. Bacteria originating from the surface may not be well adapted to the environment at depth and thus contribute little or nothing to metabolic activity there. OTUs displaying positive correlations over all positive and negative time lag points (Figure 6.7) are always more abundant in the surface than at depth. Therefore,

these are likely candidates for surface-resident organisms that reach depth only by sinking or mixing. This pattern was displayed by FlavobacteriaceaeA, *PolaribacterA* and RhodobacteraceaeA, D, E and F (Figure 6.7). In contrast, the OTUs of FlavobacteriaceaeC, PelagibacteraceaeA and B and RhodobacteraceaeC and G first occurred in the surface and were later found in higher abundances at depth, possibly an indication that sinking bacteria were metabolically-active and capable of growth at depth. The opposite occurred for *Octadecabacter*, RhodobacteraceaeB, *SUP05* and *ZA3409c*, which were brought from depth to the surface during the intrusion events and the spring bloom. *Octadecabacter* and RhodobacteraceaeB continued to grow in the surface layers, whereas *ZA3409c* and *SUP05* did not; the latter is associated with low ambient O_2 conditions, and is thus likely not viable in the oxygenated surface (Glaubnitz et al. 2009, 2010, Walsh et al. 2009). Thus, while transport mechanisms may facilitate the short-term presence of a specific species, environmental factors determine which species are able to thrive.

6.4.2.4. Diazotrophs in the Bedford Basin

The diazotrophic community changed with season and depth (Figure 6.8). As observed in the temperate ocean by Farnelid et al. (2011), sub-cluster Ip sequences dominated (596 of 927 OTUs in the rarefied sample set) and also displayed the highest read numbers (Figure 6.9). Diversity was highest in December in the surface, possibly as a result of winter-specific nutrient conditions and water column mixing. The nitrogenase has high iron requirements and an O_2-

sensitive active centre; it has been proposed that non-cyanobacterial diazotrophs find low O_2 niches in the water column such as oxygen minimum zones (OMZs; Loescher et al. 2014) or on organic molecules that are experiencing rapid degradation and a consequent drawdown of O_2 (Riemann et al. 2010). The Bedford Basin regularly displays dysoxic conditions in the winter months (< 90 µM O_2, Wright et al. 2012), which marks the initiation of anaerobic respiration and may support diazotrophic growth. In contrast, the spring sample was dominated by a single OTU from Cluster I, and thus showed the lowest diversity. This OTU could have been part of the spring bloom and thereby drastically decreasing overall diazotrophic diversity, or brought into the Bedford Basin through terrestrial water run-off, because its closest sequence in the database was identified from soil on the Tibetan Plateau (92% identity; Zhang et al. 2006).

Cyanobacterial *nifH* OTUs were detected only in the summer and autumn samples from the surface (Figure 6.9). One OTU was 92% identical to *Calothrix* sp., while all others fell into the *Candidatus* Atelocyanobacterium thalassa clade 2. Summer conditions in the Bedford Basin were much more similar to conditions that *Candidatus* A. thalassa is usually found in; nutrients were drawn down and temperatures reached 20°C (Figure 6.2; Langlois et al. 2008). It has been shown that *Candidatus* A. thalassa is a very important contributor to N_2 fixation and may contribute to the production of new fixed nitrogen species in the Bedford Basin (Martinez-Perez et al. 2016).

6.4.3. Bloom and Intrusion events

The bacterial communities in the Bedford Basin were disrupted four times during 2014. The spring and autumn bloom as well as two Atlantic Ocean water intrusions brought about rapid changes in the community. Increasing temperatures and stratification or breaking down of stratification (autumn) led to the bloom periods. Intrusion events rapidly and drastically altered environmental conditions, to which the community responded within the same week (Figure 6.2 and Supplemental Figure 19). Water intrusions came from the Atlantic Ocean, as seen in the measurements of increased salinity and decreased temperature and nutrients that are especially obvious at 60 m (Figure 6.2).

Blooms occur in the Bedford Basin every spring and autumn. Large diatoms that quickly draw down silicate as well as the other nutrients dominate the spring bloom (Li and Dickie, 2001). Smaller phytoplankton species as well as unicellular eukaryotes without silicate shells are dominant in the autumn bloom (Li and Dickie, 2001).

In 2014, the first bloom was initiated by the first oceanic water intrusion between weeks 10 and 11. The intrusion mixed the entire water column as seen in the community structure in week 11 (Figure 6.6 and Supplemental Figure 19): the Bray-Curtis similarity between the 10 and 60 m samples increased from 71% to 85% from week 10 to week 11 (Figure 6.4B), a rapid increase of RhodobacteraceaeA and *Colwellia* was observed at 60 m, which were previously only highly abundant at 1, 5 and 10 m, as also supported by the time lag analysis (Figure 6.7). Relative abundances of RhodobacteraceaeA decreased steadily at

depth in the following weeks; suggesting that deep waters is unfavourable for this OTU. The first intrusion-induced bloom is followed by two more chlorophyll spikes in the water column, which are most likely induced by warming water and increased stability of the water column (weeks 15/16 and week 19; Figure 6.2). The initial response of the bacterial community to the bloom was a radical increase in *Pseudoalteromonas*, and, to a lesser extent, *Psychrobacter* and *Phaeobacter* (weeks 12, 16 and 19). *Pseudoalteromonas* increased in relative abundance whenever chlorophyll levels spiked in the Bedford Basin, including during the spring bloom, throughout the summer, and during the autumn bloom. *Pseudoalteromonas* species have been shown to secrete various antibacterial and algicidal compounds including antibiotics, polyanionic macromolecules, small brominated compounds, toxins, agarases and proteases, all of which are very effective in causing growth inhibition or cell lysis of bacteria or algae (Vera et al. 1998; Uchida et al. 1997; Imai et al. 1995; Uchida et al.1995; Mc Carthy et al. 1994; Simidu et al. 1990; Gauthier et al. 1977; Gauthier et al. 1976; Andersen et al. 1974). Thus, it can be speculated that during its multiple blooms in the spring, *Pseudoalteromonas* was able to outcompete most bacteria and utilize nutrients from lysed phytoplankton and bacteria for rapid growth. *Phaeobacter* and *Psychrobacter*, which increased alongside *Pseudoalteromonas,* have also been shown to secrete antibacterial compounds (Li et al. 2009; Li et al. 2008; Porsby et al. 2008), giving them a competitive advantage over those without the secretion ability; they may also be less susceptible to compounds secreted by *Pseudoalteromonas* and vice versa.

Other taxa dominant in the Bedford Basin during the spring bloom period and the following weeks of spring have previously been identified as bloom-associated. These include the taxa to which the OTUs *PolaribacterA* and *B*, FlavobacteriaceaeA and B and *UlvibacterA* and *B* belong. Time lag analysis supports this conclusion (Figure 6.7). With the exception of *Phaeobacter*, these OTUs were highly positively correlated between surface and depth at the same time, suggesting that their abundances spiked simultaneously.

The autumn bloom peaked in week 38 and chlorophyll concentration remained elevated until week 42 (Figure 6.2). This is quite different to the spring bloom, which experienced much more intense spikes and plummets of chlorophyll, possibly due to the different phytoplankton taxa involved in the spring and autumn bloom (Li and Dickie 2001). Hence, the dominant bacterial community associated with the autumn bloom was also quite different from the spring bloom overall. Some OTUs were autumn bloom-specific, such as RhodobacteraceaeE and F*,* of which only RhodobacteraceaeF was found at 60 m (Figure 6.6 and Figure 6.7). As mentioned previously, the family of Rhodobacteraceae is too diverse metabolically to assign specific ecosystem function and hence, this is left here as observation only. OTUs that were also encountered during the spring bloom (*Phaeobacter*, *PolaribacterA* and *Pseudoalteromonas*), were represented in much lower relative abundances.

The second oceanic water intrusion occurred between weeks 28 and 29. The dominant species were heterogeneous across the water column at this point, and mixing by the intrusion is suggested only by total community composition that

includes the rare taxa (Bray-Curtis similarity shift from 66% to 78%, comparing 10 and 60 m at week 28 and 29; Figure 6.4B). The intrusion led to a spike in chlorophyll at 1 m depth, rather than throughout all depths as seen in the first intrusion. The Bedford Basin was extremely nitrogen limited before the intrusion event (nitrate concentration below the detection limit; Figure 6.2), and thus mixing did not induce a significant change in environmental parameter and community as was observed in the spring.

Overall, the magnitude of the microbial community response to environmental change varies with the specific change and, perhaps, the availability of nutrients. Significant change is seen when ample nutrients are available (first intrusion and spring bloom), whereas the change is much less pronounced when the community appears nutrient limited (second intrusion and autumn bloom).

Intrusion and bloom events highlight that studies lacking a temporal aspect – as it is often the case during research cruises due to resource limitation - can only provide a snapshot of the community; if this is captured during a moment of disturbance, it could ultimately lead to a skewed picture of prevailing community compositions. It is also clear that seasonality alters the predominant taxa and extrapolation from a single measurement could easily lead to false conclusions.

6.4.4. Challenges

Through the advancement of the Illumina techniques, it has become possible to increase read lengths from <150 bp to <500 bp. The particular read length for this

study was approximately 400 bp, and while this has already increased taxonomic resolution, it is not sufficient to classify all OTUs to the species or even genus level, a task which also depends on sequence availability in databases. Compared to the 2014 study of El-Swais et al, read length was increased via a different variable region (V6-V8 rather than V5) and a higher clustering threshold (97% rather than 90%); both of these would affect OTU classification, implying that caution is crucial when making direct comparisons. We also used a larger sampling volume (500 mL versus 2 mL), did not use a chemical preservative and expanded sampling to four depths, which altogether has given us a very detailed insight into the bacterial community in the Bedford Basin. However, all studies employing high-throughput sequencing technologies face certain challenges, including the introduction of PCR amplification bias based on the choice of primers or randomness of amplification (Engelbrektson et al. 2010). The challenge of amplification randomness was addressed by running PCRs with several template dilutions. Results obtained from PCR amplified samples should always be interpreted with the knowledge that some sequences may have been amplified preferentially, and that abundances are relative rather than quantitative.

6.5. Conclusion and outlook

This study presents a highly resolved time-series in a temperate ocean inlet. It observed that the bacterial community transitions throughout the year in accordance with the changing environment, and returns at the end of the year, to a community similar to that at the outset. The main drivers of community similarity

were identified as temperature, nutrient concentration, O_2 concentration, and the presence of phytoplankton. Abrupt changes in the environment led to a rapid change in microbial community composition. Since environmental conditions will be altered with climate change, we predict that the bacterial community in the Bedford Basin will respond to those changes rapidly. We also showed seasonal variability of the extremely diverse diazotrophic community.

Altogether we believe that, due to its geographic location and the already existing extensive monitoring program, the Bedford Basin provides an excellent model for the study of variations in microbial community structure over time: seasonal, annual and decadal. In the long-term, we aim to establish a microbial time series in the Bedford Basin that can be connected with weekly measurements of physical and chemical parameters by the Bedford Institute of Oceanography, which will enable us to monitor and predict marine microbial communities in temperate coastal regions as the oceans change under anthropogenic influence.

6.6. Acknowledgements

Jenni-Marie Ratten's work was supported by the NSERC CREATE Transatlantic Ocean System Science and Technology (TOSST) grant, the Stiftung für Kanadastudien (SKS), the Deutscher Akademischer Austausch Dienst (DAAD) and Julie LaRoche's NSERC Discovery research grant.

6.7. References

Alvarez, R. and Vera, J., 1998. Identification of a marine agarolytic *Pseudoalteromonas* isolate and characterization of its extracellular agarase. Applied and Environmental Microbiology, 64(11), pp.4378-4383.

Anantharaman, K., Breier, J.A., Sheik, C.S. and Dick, G.J., 2013. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proceedings of the National Academy of Sciences, 110(1), pp.330-335.

Andersen, R.J., Wolfe, M.S. and Faulkner, D.J., 1974. Autotoxic antibiotic production by a marine *Chromobacterium*. Marine Biology, 27(4), pp.281-285.

Aota, Y. and Nakajima, H., 2001. Mutualistic relationships between phytoplankton and bacteria caused by carbon excretion from phytoplankton. Ecological research, 16(2), pp.289-299.

Arrigo, K.R., 2005. Marine microorganisms and global nutrient cycles. Nature, 437(7057), pp.349-355.

Azam, F., 1998. Microbial control of oceanic carbon flux: the plot thickens. Science, 280(5364), pp.694-696.

Baek, K., Jo, H., Choi, A., Kang, I. and Cho, J.C., 2014. *Ulvibacter marinus* sp. nov., isolated from coastal seawater. International journal of systematic and evolutionary microbiology, 64(6), pp.2041-2046.

Bowman, J.S., Rasmussen, S., Blom, N., Deming, J.W., Rysgaard, S. and Sicheritz-Ponten, T., 2012. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. The ISME journal, 6(1), pp.11-20.

Buchan, A., González, J.M. and Moran, M.A., 2005. Overview of the marine Roseobacter lineage. Applied and environmental microbiology, 71(10), pp.5665-5677.

Bunse, C., Lundin, D., Karlsson, C.M., Akram, N., Vila-Costa, M., Palovaara, J., Svensson, L., Holmfeldt, K., González, J.M., Calvo, E. and Pelejero, C., 2016. Response of marine bacterioplankton pH homeostasis gene expression to elevated CO2. Nature Climate Change, 6(5), pp.483-487.

Calbet, A. and Landry, M.R., 2004. Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. Limnology and Oceanography, 49(1), pp.51-57.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I. and Huttley, G.A., 2010a. QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7(5), pp.335-336.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R., 2010b. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics, 26(2), pp.266-267.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), pp.540-552.

Cho, J.C. and Giovannoni, S.J., 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria. Applied and Environmental Microbiology, 70(1), pp.432-440.

Chow, C.E.T., Sachdeva, R., Cram, J.A., Steele, J.A., Needham, D.M., Patel, A., Parada, A.E. and Fuhrman, J.A., 2013. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. The ISME journal, 7(12), pp.2259-2273.

Clarke, K.R. and Gorley, R.N., 2006. PRIMER version 6: user manual/tutorial. PRIMER-E, Plymouth, UK, 192.

Codispoti, L.A., 2007. An oceanic fixed nitrogen sink exceeding 400 Tg N a-1 vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences, 4(2), pp.233-253.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W. and Yoshinari, T., 2001. The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene?. Scientia Marina, 65(S2), pp.85-105.

Comeau, A.M., Douglas, G.M. and Langille, M.G.I., 2016. A custom and streamlined workflow for microbiome research. mSystems, submitted.

Comeau, A.M., Douglas, G.M. and Langille, M.G., 2017. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. mSystems, 2(1), pp.e00127-16.

Comeau, A.M., Li, W.K., Tremblay, J.É., Carmack, E.C. and Lovejoy, C., 2011. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLoS One, 6(11), p.e27492.

Cram, J.A., Chow, C.E.T., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A. and Fuhrman, J.A., 2015b. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. The ISME journal, 9(3), pp.563-580.

Cram, J.A., Xia, L.C., Needham, D.M., Sachdeva, R., Sun, F. and Fuhrman, J.A., 2015a. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. The ISME journal, 9(12), pp.2573-2586.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I. and Carmichael, M., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science, 348(6237), p.1261605.

Diaz, R.J. and Rosenberg, R., 2008. Spreading dead zones and consequences for marine ecosystems. science, 321(5891), pp.926-929.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14(6), pp.927-930.

Doney, S.C., Ruckelshaus, M., Duffy, J.E., Barry, J.P., Chan, F., English, C.A., Galindo, H.M., Grebmeier, J.M., Hollowed, A.B., Knowlton, N. and Polovina, J., 2012. Climate change impacts on marine ecosystems. Marine Science, 4.

Dong, S., Yang, J., Zhang, X.Y., Shi, M., Song, X.Y., Chen, X.L. and Zhang, Y.Z., 2012. Cultivable alginate lyase-excreting bacteria associated with the arctic brown alga Laminaria. Marine drugs, 10(11), pp.2481-2491.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27(16), pp.2194-2200.

El-Swais, H., Dunn, K.A., Bielawski, J.P., Li, W.K. and Walsh, D.A., 2015. Seasonal assemblages and short-lived blooms in coastal north-west Atlantic Ocean bacterioplankton. Environmental microbiology, 17(10), pp.3642-3661.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H. and Hugenholtz, P., 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. The ISME journal, 4(5), pp.642-647.

Falkowski, P.G., Fenchel, T. and Delong, E.F., 2008. The microbial engines that drive Earth's biogeochemical cycles. science, 320(5879), pp.1034-1039.

Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å. and Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PLoS One, 6(4), p.e19223.

Fernández-Méndez, M., Turk-Kubo, K.A., Buttigieg, P.L., Rapp, J.Z., Krumpen, T., Zehr, J.P. and Boetius, A., 2016. Diazotroph Diversity in the Sea Ice, Melt Ponds, and Surface Waters of the Eurasian Basin of the Central Arctic Ocean. Frontiers in Microbiology, 7.

Forth, M., Liljebladh, B., Stigebrandt, A., Hall, P.O. and Treusch, A.H., 2015. Effects of ecological engineered oxygenation on the bacterial community structure in an anoxic fjord in western Sweden. The ISME journal, 9(3), pp.656-669.

Fouilland, E., Tolosa, I., Bonnet, D., Bouvier, C., Bouvier, T., Bouvy, M., Got, P., Le Floc'h, E., Mostajir, B., Roques, C. and Sempéré, R., 2014. Bacterial carbon dependence on freshly produced phytoplankton exudates under different nutrient availability and grazing pressure conditions in coastal marine waters. FEMS microbiology ecology, 87(3), pp.757-769.

Friedman, J. and Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol, 8(9), p.e1002687.

Fu, Y., Keats, K.F., Rivkin, R.B. and Lang, A.S., 2013. Water mass and depth determine the distribution and diversity of Rhodobacterales in an Arctic marine system. FEMS microbiology ecology, 84(3), pp.564-576.

Fuhrman, J.A., Cram, J.A. and Needham, D.M., 2015. Marine microbial community dynamics and their ecological interpretation. Nature Reviews Microbiology, 13(3), pp.133-146.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V. and Naeem, S., 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences, 103(35), pp.13104-13109.

Fulweiler, R.W., Brown, S.M., Nixon, S.W. and Jenkins, B.D., 2013. Evidence and a conceptual model for the co?occurrence of nitrogen fixation and denitrification in heterotrophic marine sediments. Marine Ecology Progress Series, 482, pp.57-68.

Gauthier, M.J., 1977. Alteromonas citrea, a new Gram-negative, yellow-pigmented species from seawater. International Journal of Systematic and Evolutionary Microbiology, 27(4), pp.349-354.

Gauthier, M.J. and Flatau, G.N., 1976. Antibacterial activity of marine violet-pigmented Alteromonas with special reference to the production of brominated compounds. Canadian Journal of Microbiology, 22(11), pp.1612-1619.

Georges, A.A., El-Swais, H., Craig, S.E., Li, W.K. and Walsh, D.A., 2014. Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. ISME J, 8(6), pp.1301-1313.

Ghiglione, J.F., Galand, P.E., Pommier, T., Pedrós-Alió, C., Maas, E.W., Bakker, K., Bertilson, S., Kirchman, D.L., Lovejoy, C., Yager, P.L. and Murray, A.E., 2012. Pole-to-pole biogeography of surface and deep marine bacterial communities. Proceedings of the National Academy of Sciences, 109(43), pp.17633-17638.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I. and Somerfield, P., 2012. Defining seasonal marine microbial community dynamics. The ISME journal, 6(2), pp.298-308.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M. and Rappé, M.S., 2005. Genome streamlining in a cosmopolitan oceanic bacterium. science, 309(5738), pp.1242-1245.

Giovannoni, S.J. and Vergin, K.L., 2012. Seasonality in ocean microbial communities. Science, 335(6069), pp.671-676.

Glaubitz, S., Lueders, T., Abraham, W.R., Jost, G., Jürgens, K. and Labrenz, M., 2009. 13C-isotope analyses reveal that chemolithoautotrophic Gamma-and Epsilonproteobacteria feed a microbial food web in a pelagic redoxcline of the central Baltic Sea. Environmental microbiology, 11(2), pp.326-337.

Glaubitz, S., Labrenz, M., Jost, G. and Jürgens, K., 2010. Diversity of active chemolithoautotrophic prokaryotes in the sulfidic zone of a Black Sea pelagic redoxcline as determined by rRNA-based stable isotope probing. FEMS microbiology ecology, 74(1), pp.32-41.

Gosink, J.J., Woese, C.R. and Staley, J.T., 1998. Polaribacter gen. nov., with three new species, P. irgensii sp. nov., P. franzmannii sp. nov. and P. filamentus sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of ‘Flectobacillus glomeratus’ as Polaribacter glomeratus comb. nov. International Journal of Systematic and Evolutionary Microbiology, 48(1), pp.223-235.

Gram, L., Melchiorsen, J. and Bruhn, J.B., 2010. Antibacterial activity of marine culturable bacteria collected from a global sampling of ocean surface waters and surface swabs of marine organisms. Marine biotechnology, 12(4), pp.439-451.

Grice, E.A. and Segre, J.A., 2011. The skin microbiome. Nature Reviews Microbiology, 9(4), pp.244-253.

Gruber, N., 2008. The marine nitrogen cycle: overview and challenges. Nitrogen in the marine environment, pp.1-50.

Hamersley, M.R., Lavik, G., Woebken, D., Rattray, J.E., Lam, P., Hopmans, E.C., Damsté, J.S.S., Krüger, S., Graco, M., Gutiérrez, D. and Kuypers, M.M., 2007. Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. Limnology and Oceanography, 52(3), pp.923-933.

Hellou, J., Steller, S., Zitko, V., Leonard, J., King, T., Milligan, T.G. and Yeats, P., 2002. Distribution of PACs in surficial sediments and bioavailability to mussels, Mytilus edulis of Halifax Harbour. Marine environmental research, 53(4), pp.357-379.

Huete-Stauffer, T.M., Arandia-Gorostidi, N., Díaz-Pérez, L. and Moran, X.A.G., 2015. Temperature dependences of growth rates and carrying capacities of marine bacteria depart from metabolic theoretical predictions. FEMS microbiology ecology, 91(10), p.fiv111.

Huston, A.L., 2003. Bacterial adaptation to the cold: in Sim aotivities of extracellular (Doctoral dissertation, University of Washington).

Huston, A.L., Krieger-Brockett, B.B. and Deming, J.W., 2000. Remarkably low temperature optima for extracellular enzyme activity from Arctic bacteria and sea ice. Environmental Microbiology, 2(4), pp.383-388.

Huston, A.L., Methe, B. and Deming, J.W., 2004. Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile Colwellia psychrerythraea strain 34H. Applied and environmental microbiology, 70(6), pp.3321-3328.

Hutchins, D.A., Mulholland, M.R. and Fu, F., 2009. Nutrient cycles and marine microbes in a CO2-enriched ocean. Oceanography.

Imai, I., Ishida, Y., Sakaguchi, K. and Hata, Y., 1995. Algicidal marine bacteria isolated from northern Hiroshima Bay, Japan. Fisheries science, 61(4), pp.628-636.

Ingall, E. and Jahnke, R., 1994. Evidence for enhanced phosphorus regeneration from marine sediments overlain by oxygen depleted waters. Geochimica et Cosmochimica Acta, 58(11), pp.2571-2575.

Jacoby, W.G., 2000. Loess:: a nonparametric, graphical tool for depicting relationships between variables. Electoral Studies, 19(4), pp.577-613.

Joint, I., Tait, K., Callow, M.E., Callow, J.A., Milton, D., Williams, P. and Cámara, M., 2002. Cell-to-cell communication across the prokaryote-eukaryote boundary. Science, 298(5596), pp.1207-1207.

Karl, D.M. and Church, M.J., 2014. Microbial oceanography and the Hawaii Ocean Time-series programme. Nature Reviews Microbiology, 12(10), pp.699-713.

Kirchman, D.L., 2000. Uptake and regeneration of inorganic nutrients by marine heterotrophic bacteria. Microbial ecology of the oceans, pp.261-288.

Klindworth, A., Mann, A.J., Huang, S., Wichels, A., Quast, C., Waldmann, J., Teeling, H. and Glöckner, F.O., 2014. Diversity and activity of marine

bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. Marine genomics, 18, pp.185-192.

Kopylova, E., Noé, L. and Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics, 28(24), pp.3211-3217.

Kujawinski, E.B., Longnecker, K., Barott, K.L., Weber, R.J.M. and Kido Soule, M.C., 2016. Microbial Community Structure Affects Marine Dissolved Organic Matter Composition. Front. Mar. Sci, 3, p.45.

Li, W. K., 2014. The state of phytoplankton and bacterioplankton at the Compass Buoy Station: Bedford Basin Monitoring Program 1992-2013. Canadian Technical Report of Hydrography and Ocean Sciences 304.

Li, W.K. and Glen Harrison, W., 2008. Propagation of an atmospheric climate signal to phytoplankton in a small marine basin. Limnology and Oceanography, 53(5), pp.1734-1745.

Li, W.K., Harrison, W.G. and Head, E.J., 2006. Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. Proceedings of the Royal Society of London B: Biological Sciences, 273(1596), pp.1953-1960.

Li, W.K.W. and Dickie, P.M., 2001. Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. Cytometry, 44(3), pp.236-246.

Li, H., Lee, B.C., Kim, T.S., Bae, K.S., Hong, J.K., Choi, S.H., Bao, B. and Jung, J.H., 2008. Bioactive cyclic dipeptides from a marine sponge-associated bacterium, Psychrobacter sp. Biomolecules and Therapeutics, 16(4), pp.356-363.

Li, H., Shinde, P.B., Lee, H.J., Yoo, E.S., Lee, C.O., Hong, J., Choi, S.H. and Jung, J.H., 2009. Bile acid derivatives from a sponge-associated bacterium Psychrobacter sp. Archives of pharmacal research, 32(6), pp.857-862.

Loescher, C.R., Großkopf, T., Desai, F.D., Gill, D., Schunck, H., Croot, P.L., Schlosser, C., Neulinger, S.C., Pinnow, N., Lavik, G. and Kuypers, M.M., 2014. Facets of diazotrophy in the oxygen minimum zone waters off Peru. The ISME journal, 8(11), pp.2180-2192.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. and Knight, R., 2012. Diversity, stability and resilience of the human gut microbiota. Nature, 489(7415), pp.220-230.

Lucas, J., Wichels, A., Teeling, H., Chafee, M., Scharfe, M. and Gerdts, G., 2015. Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. FEMS microbiology ecology, 91(9), p.fiv099.

Lynch, M.D. and Neufeld, J.D., 2015. Ecology and exploration of the rare biosphere. Nature Reviews Microbiology, 13(4), pp.217-229.

Marx, J.G., Carpenter, S.D. and Deming, J.W., 2009. Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium Colwellia psychrerythraea strain 34H under extreme conditions This article is one of a selection of papers in the Special Issue on

Polar and Alpine Microbiology. Canadian journal of microbiology, 55(1), pp.63-72.

Mason, O.U., Han, J., Woyke, T. and Jansson, J.K., 2015. Single-cell genomics reveals features of a Colwellia species that was dominant during the Deepwater Horizon oil spill. The metabolic pathways and environmental controls of hydrocarbon biodegradation in marine ecosystems, p.29.

Mason, O.U., 2014. Metagenomics, metatranscriptomics and single cell genomics reveal functional response of active Oceanospirillales to Gulf oil spill. Journal of Microbial Ecology (The ISME).

McCarthy, S.A., Johnson, R.M. and Kakimoto, D., 1994. Characterization of an antibiotic produced by Alteromonas luteoviolacea Gauthier 1982, 85 isolated from Kinko Bay, Japan. Journal of applied bacteriology, 77(4), pp.426-432.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. and Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal, 6(3), pp.610-618.

Mercier, C., Boyer, F., Bonin, A. and Coissac, E., 2013. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. Programs Abstr SeqBio, pp.27-9.

Methé, B.A., Nelson, K.E., Deming, J.W., Momen, B., Melamud, E., Zhang, X., Moult, J., Madupu, R., Nelson, W.C., Dodson, R.J. and Brinkac, L.M., 2005. The psychrophilic lifestyle as revealed by the genome sequence of Colwellia psychrerythraea 34H through genomic and proteomic analyses. Proceedings of the National Academy of Sciences of the United States of America, 102(31), pp.10913-10918.

Moeseneder, M.M., Arrieta, J.M. and Herndl, G.J., 2005. A comparison of DNA- and RNA-based clone libraries from the same marine bacterioplankton community. FEMS microbiology ecology, 51(3), pp.341-352.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A. and Giovannoni, S.J., 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature, 420(6917), pp.806-810.

Nedashkovskaya, O.I., Kim, S.B., Han, S.K., Rhee, M.S., Lysenko, A.M., Falsen, E., Frolova, G.M., Mikhailov, V.V. and Bae, K.S., 2004. Ulvibacter litoralis gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the green alga Ulva fenestrata. International journal of systematic and evolutionary microbiology, 54(1), pp.119-123.

Nedashkovskaya, O.I., Kim, S.B., Vancanneyt, M., Snauwaert, C., Lysenko, A.M., Rohde, M., Frolova, G.M., Zhukova, N.V., Mikhailov, V.V., Bae, K.S. and Oh, H.W., 2006. Formosa agariphila sp. nov., a budding bacterium of the family Flavobacteriaceae isolated from marine environments, and emended description of the genus Formosa. International journal of systematic and evolutionary microbiology, 56(1), pp.161-167.

Neuwirth, E., 2011. RColorBrewer: ColorBrewer palettes. R package version, 1(5).

Pomeroy, L.R. and Wiebe, W.J., 2001. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. Aquatic Microbial Ecology, 23(2), pp.187-204.

Pommier, T., Pinhassi, J. and Hagström, Å., 2005. Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. Aquatic Microbial Ecology, 41(1), pp.79-89.

Porsby, C.H., Nielsen, K.F. and Gram, L., 2008. Phaeobacter and Ruegeria species of the Roseobacter clade colonize separate niches in a Danish turbot (Scophthalmus maximus)-rearing farm and antagonize Vibrio anguillarum under different growth conditions. Applied and environmental microbiology, 74(23), pp.7356-7364.

Price, M.N., Dehal, P.S. and Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one, 5(3), p.e9490.

Price, P.B. and Sowers, T., 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. Proceedings of the National Academy of Sciences of the United States of America, 101(13), pp.4631-4636.

Raes, J., Letunic, I., Yamada, T., Jensen, L.J. and Bork, P., 2011. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. Molecular systems biology, 7(1), p.473.

Riemann, L., Farnelid, H. and Steward, G.F., 2010. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. Aquatic Microbial Ecology, 61(3), pp.235-247.

Schlitzer, R., 2015. Ocean Data View. 2012.

Simidu, U., Kita-Tsukamoto, K.U.M.I.K.O., YASUMOTO, T. and YOTSU, M., 1990. Taxonomy of four marine bacterial strains that produce tetrodotoxin. International Journal of Systematic and Evolutionary Microbiology, 40(4), pp.331-336.

Spietz, R.L., Williams, C.M., Rocap, G. and Horner-Devine, M.C., 2015. A Dissolved Oxygen Threshold for Shifts in Bacterial Community Structure in a Seasonally Hypoxic Estuary. PloS one, 10(8), p.e0135731.

Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.E.T., Sachdeva, R., Jones, A.C., Schwalbach, M.S. and Rose, J.M., 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME journal, 5(9), pp.1414-1425.

Stingl, U., Desiderio, R.A., Cho, J.C., Vergin, K.L. and Giovannoni, S.J., 2007. The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. Applied and environmental microbiology, 73(7), pp.2290-2296.

Stoffer, D., 2012. astsa: Applied Statistical Time Series Analysis. R package version, 1.

Stramma, L., Johnson, G.C., Sprintall, J. and Mohrholz, V., 2008. Expanding oxygen-minimum zones in the tropical oceans. science, 320(5876), pp.655-658.

Strom, S.L., 2008. Microbial ecology of ocean biogeochemistry: a community perspective. Science, 320(5879), pp.1043-1045.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. and Cornejo-Castillo, F.M., 2015. Structure and function of the global ocean microbiome. Science, 348(6237), p.1261359.

Sunamura, M., Higashi, Y., Miyako, C., Ishibashi, J.I. and Maruyama, A., 2004. Two bacteria phylotypes are predominant in the Suiyo Seamount hydrothermal plume. Applied and Environmental Microbiology, 70(2), pp.1190-1198.

Suttle, C.A., 2007. Marine viruses—major players in the global ecosystem. Nature Reviews Microbiology, 5(10), pp.801-812.

Team, R.C., 2015. R: A language and environment for statistical computing.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., Kassabgy, M., Huang, S., Mann, A.J., Waldmann, J. and Weber, M., 2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science, 336(6081), pp.608-611.

Turnbaugh, P.J., Biomolecules, S.B.D. and Roscoff, F., 2011. Environmental and gut bacteroidetes: the food connection. Human health and disease in a microbial world, p.96.

Uchida, M., Nakata, K. and Maeda, M., 1997. Conversion of Ulva fronds to a hatchery diet for Artemia nauplii utilizing the degrading and attaching abilities of Pseudoalteromonas espejiana. Journal of applied phycology, 9(6), pp.541-549.

Uchida, M., Nakayama, A. and Abe, S., 1995. Distribution and characterization of bacteria capable of decomposing brown algae fronds in waters associated with Laminaria vegetation. Fisheries science, 61(1), pp.117-120.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P.C., Colwell, R.R., Höfle, M.G. and Pruzzo, C., 2012. Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. The ISME journal, 6(1), pp.21-30.

Vollmers, J., Voget, S., Dietrich, S., Gollnow, K., Smits, M., Meyer, K., Brinkhoff, T., Simon, M. and Daniel, R., 2013. Poles apart: Arctic and Antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. PLoS One, 8(5), p.e63422.

Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, S.G., Tortell, P.D. and Hallam, S.J., 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science, 326(5952), pp.578-582.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology, 73(16), pp.5261-5267.

Warnes, M.G.R., 2016. Package ‘gplots’.

Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J. and Birol, I., 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. GigaScience, 4(1), p.35.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R. and Ley, R.E., 2012. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME journal, 6(1), pp.94-103.

Wickham, H. and Chang, W., 2015. Ggplot2: An implementation of the grammar of graphics, Version 1.0. 1.

Wright, J.J., Konwar, K.M. and Hallam, S.J., 2012. Microbial ecology of expanding oxygen minimum zones. Nature Reviews Microbiology, 10(6), pp.381-394.

Yoon, J.H., Kang, S.J. and Oh, T.K., 2006. Polaribacter dokdonensis sp. nov., isolated from seawater. International journal of systematic and evolutionary microbiology, 56(6), pp.1251-1255.

Zehr, J.P. and Kudela, R.M., 2011. Nitrogen cycle of the open ocean: from genes to ecosystems. Annual Review of Marine Science, 3, pp.197-225.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), pp.614-620.

Zhang, Y., Li, D., Wang, H., Xiao, Q. and Liu, X., 2006. Molecular diversity of nitrogen-fixing bacteria from the Tibetan Plateau, China. FEMS microbiology letters, 260(2), pp.134-142.

CHAPTER 7: CONCLUSION

Climate change will have a profound effect on oceanic cycles and systems. Predictions include increasing temperatures with longer-lasting stratification periods, acidification, and the expansion of oxygen minimum zones (Stocker et al. 2014). As microbial communities are starting to adapt to a changing ocean, their responses will influence overall effects on the greater oceanic ecosystem networks (Fuhrman et al. 2015; Reusch 2014; Vezzulli et al. 2012; Wright et al. 2012). In the vast areas of surface ocean that are limited in nitrogen, primary productivity is controlled by the remineralization of nitrogen species and by newly fixed N_2 from diazotrophs (Moore et al. 2013). Consequently, permanent removal of CO_2 from the atmosphere into the deep ocean is tightly linked to N_2 fixation (Azam 1998). This underlines the importance of understanding the distribution, diversity and metabolism of diazotrophs to predict their responses to climate change and the role they may play in the complex network of feedback systems.

A review of recent findings combined with an in-depth investigation of the Atlantic Ocean using Illumina paired-end sequencing showed that non-cyanobacterial *nifH* sequences are widely distributed throughout all oceanic environments and that they make up a large proportion of the diazotrophic community in the temperate ocean, below the photic zone, and in low O_2 regimes (Chapters 3 and 4). This stresses that non-cyanobacterial organisms should not be neglected when studying the ocean's diazotrophs. Compared to the cyanobacterial cluster, non-cyanobacterial diazotrophs contain many more unique *nifH* sequences of which the majority remain unidentified; this demonstrates the urgent need to

further investigate these organisms using quantitative rather than relative methods, and most importantly, the metabolic role they play in their specific microbial communities as well as in the global nitrogen cycle (Chapter 3). Other high-throughput sequencing methods such as targeted metagenomics and metatranscriptomics could play an important role in creating a global picture of non-cyanobacterial diazotrophs in a culture-independent way. Attaining more insight into the metabolic potential of these organisms would lead to a better understanding of their life-styles, distribution patterns, and might allow for improved cultivation techniques.

Cultivation allows for in-depth examinations of diazotrophic metabolism, specifically their N_2 fixation rates. Only a few heterotrophic organisms have been cultivated so far, but current isolates already show great diversity in optimal growth and N_2 fixation conditions (Bentzon-Tilia et al 2015, 2014). We isolated a heterotrophic diazotroph from the Bedford Basin, which was shown to be widespread throughout the temperate North Atlantic (Chapter 5). The conditions under which it performs N_2 fixation (e.g. O_2 concentrations, carbon sources and DIN concentrations) as well as N_2 fixation rates still need to be determined. Nonetheless, based on estimated qPCR abundances and distribution, and its metabolism as inferred by genome sequencing, we predict that its role in microbial community far exceeds that of N_2 fixation and that it has the potential to be an important player in the temperate diazotrophic community.

To gain a broader understanding of the metabolism of non-cyanobacterial diazotrophs, metabolic pathways concerning nutrient cycling were analyzed in

132 reference genomes (112 genomes of non-cyanobacterial species) that aligned closely with marine uncultured *nifH* sequences (Chapter 4). The metabolism of these diazotrophs was very diverse, especially that of proteobacterial reference genomes, providing a possible explanation for their wide distribution throughout aquatic environments. Still to be determined is the activity of these organisms in the ocean. This could again be addressed using metagenomics, metaproteomics or more targeted approaches such as probe-based quantification used in NanoString assays or qPCR (Poong et al. 2017). To predict diazotrophic response to a changing ocean, these findings will then have to be paralleled with environmental parameters to allow for changes in diazotrophs' behaviour to be inferred as these parameters change. Interdisciplinary research will be key to combine environmental parameters with biological measurements and metabolic modeling studies.

I set out to correlate abundance variability of seven diazotrophic phylotypes with over 100 hydrographic and environmental parameters (Chapter 2). That way I confirmed the importance of aeolian dust deposition in the tropical eastern Atlantic ecosystem for diazotrophs and that distribution at depth was associated with high PO_4^{3-} concentrations as found in water masses from either the oxygen minimum zone near the African Coast or the Gulf Stream on the western side of the Atlantic.

One major pitfall of spatial microbial studies is the lack of the time dimension. A snap shot obtained during sampling may result in a skewed picture of a microbial community in the case of recent disturbances, because microbial communities

respond rapidly to change. Additionally, temperate time-series of microbial communities can act as models for climate change as they progress through cycles of mixed winter conditions to stratified summer periods. This thesis showed a varying microbial and very diverse diazotrophic community in the Bedford Basin over the period of one year. Changes in the most abundant OTUs occurred gradually through the seasons, but sometimes were abrupt in response to extreme environmental shifts. The main environmental drivers of community changes were identified to be temperature, nutrient concentration, O_2 concentration and the presence of phytoplankton. Weekly sampling provided a good sampling frequency to observe gradual changes in the dominant microbial community. However, in specific circumstances such as bloom periods, a more frequent sampling regime should be considered to determine the exact progression of the community change. Additionally, analyses of rare taxa, metagenomics and metatranscriptomics should be explored further to create a complete picture of the microbial community and its function in the Bedford Basin.

Overall, this thesis has expanded on the knowledge of the non-cyanobacterial diazotrophs in the North Atlantic Ocean by using cultivating and culture-independent methods. Going forward, it will be crucial to continue characterizing the role, the metabolism and the N_2 fixation rates of non-cyanobacterial diazotrophs throughout the oceans. Diazotrophic responses to climate change will have a major impact on primary producers and consequently along the food web down to the amount of carbon exported into the deep sea. Understanding

these mechanisms might allow for more accurate projections into the future of global nutrient cycles.

REFERENCES

Abell, J., Emerson, S. and Renaud, P., 2000. Distributions of TOP, TON and TOC in the North Pacific subtropical gyre: Implications for nutrient supply in the surface ocean and remineralization in the upper thermocline. Journal of Marine Research, 58(2), pp.203-222.

Adachi, K., Nakatani, M. and Mochida, H., 2002. Isolation of an endophytic diazotroph, Klebsiella oxytoca, from sweet potato stems in Japan. Soil science and plant nutrition, 48(6), pp.889-895.

Akaraonye, E., Keshavarz, T. and Roy, I., 2010. Production of polyhydroxyalkanoates: the future green materials of choice. Journal of Chemical Technology and Biotechnology, 85(6), pp.732-743.

Alber, B.E., Spanheimer, R., Ebenau-Jehle, C. and Fuchs, G., 2006. Study of an alternate glyoxylate cycle for acetate assimilation by Rhodobacter sphaeroides. Molecular microbiology, 61(2), pp.297-309.

Altabet, M.A., 2006. Constraints on oceanic N balance/imbalance from sedimentary 15 N records. Biogeosciences Discussions, 3(4), pp.1121-1155.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

Alvarez, R. and Vera, J., 1998. Identification of a Marine Agarolytic Pseudoalteromonas Isolate and Characterization of Its. Applied and Environmental Microbiology, 64(11), pp.4378-4383.

Anantharaman, K., Breier, J.A., Sheik, C.S. and Dick, G.J., 2013. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proceedings of the National Academy of Sciences, 110(1), pp.330-335.

Andersen, R.J., Wolfe, M.S. and Faulkner, D.J., 1974. Autotoxic antibiotic production by a marine Chromobacterium. Marine Biology, 27(4), pp.281-285.

Andersson, B., Sundbäck, K., Hellman, M., Hallin, S. and Alsterberg, C., 2014. Nitrogen fixation in shallow-water sediments: Spatial distribution and controlling factors. Limnology and Oceanography, 59(6), pp.1932-1944.

Anzaldi, L.L. and Skaar, E.P., 2010. Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. Infection and immunity, 78(12), pp.4977-4989.

Aota, Y. and Nakajima, H., 2001. Mutualistic relationships between phytoplankton and bacteria caused by carbon excretion from phytoplankton. Ecological research, 16(2), pp.289-299.

Arrigo, K.R., 2005. Marine microorganisms and global nutrient cycles. Nature, 437(7057), pp.349-355.

Azam, F., 1998. Microbial control of oceanic carbon flux: the plot thickens. Science, 280(5364), pp.694-696.

Baek, K., Jo, H., Choi, A., Kang, I. and Cho, J.C., 2014. Ulvibacter marinus sp. nov., isolated from coastal seawater. International journal of systematic and evolutionary microbiology, 64(6), pp.2041-2046.

Baker, A., Jickells, T., Witt, M., Linge, K., 2006. Trends in the solubility of iron, aluminium, manganese and phosphorus in aerosol collected over the Atlantic Ocean. Marine Chemistry, 98(1), pp.43-58.

Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stöckel, J., Liberton, M., Min, H., Sherman, L.A. and Pakrasi, H.B., 2011. Novel metabolic attributes of the genus Cyanothece, comprising a group of unicellular nitrogen-fixing cyanobacteria. MBio, 2(5), pp.e00214-11.

Béja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L. and Spudich, E.N., 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science, 289(5486), pp.1902-1906.

Benavides, M. and Voss, M., 2015. Five decades of N_2 fixation research in the North Atlantic Ocean. Frontiers in Marine Science, 2, p.40.

Benavides, M., Bonnet, S., Hernández, N., Martínez-Pérez, A.M., Nieto-Cid, M., Álvarez-Salgado, X.A., Baños, I., Montero, M.F., Mazuecos, I.P., Gasol, J.M. and Osterholz, H., 2016. Basin-wide N_2 fixation in the deep waters of the Mediterranean Sea. Global Biogeochemical Cycles, 30(6), pp.952-961.

Benavides, M., Luo, Y.W., Doney, S.C., Anderson, L.A., Bode, A., Bonnet, S., Boström, K.H., Böttjer, D., Capone, D.G., Carpenter, E.J. and Chen, Y.L., 2012. Database of diazotrophs in global ocean: abundances, biomass and nitrogen fixation rates.

Bentzon-Tilia, M., Farnelid, H., Jürgens, K. and Riemann, L., 2014. Cultivation and isolation of N_2-fixing bacteria from suboxic waters in the Baltic Sea. FEMS microbiology ecology, 88(2), pp.358-371.

Bentzon-Tilia, M., Traving, S.J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J.L., Markager, S. and Riemann, L., 2015. Significant N_2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. The ISME journal, 9(2), pp.273-285.

Berelson, W.M., 1991. The flushing of two deep-sea basins, southern California borderland. Limnology and Oceanography, 36(6), pp.1150-1166.

Bergquist, B.A. and Boyle, E.A., 2006a. Dissolved iron in the tropical and subtropical Atlantic Ocean. Global Biogeochemical Cycles, 20(1).

Bergquist, B.A. and Boyle, E.A., 2006b. Iron isotopes in the Amazon River system: Weathering and transport signatures. Earth and Planetary Science Letters, 248(1), pp.54-68.

Berman-Frank, I., Cullen, J.T., Shaked, Y., Sherrell, R.M. and Falkowski, P.G., 2001. Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. Limnology and Oceanography, 46(6), pp.1249-1260.

Berman-Frank, I., Lundgren, P. and Falkowski, P., 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Research in microbiology, 154(3), pp.157-164.

Berman-Frank, I., Quigg, A., Finkel, Z.V., Irwin, A.J. and Haramaty, L., 2007. Nitrogen-fixation strategies and Fe requirements in cyanobacteria. Limnology and Oceanography, 52(5), pp.2260-2269.

Bertics, V.J., Löscher, C.R., Salonen, I., Dale, A.W., Gier, J., Schmitz, R.A. and Treude, T., 2013. Occurrence of benthic microbial nitrogen fixation coupled to sulfate reduction in the seasonally hypoxic Eckernförde Bay, Baltic Sea. Biogeosciences, 10(3), pp.1243-1258.

Bethoux, J.P., Morin, P., Madec, C. and Gentili, B., 1992. Phosphorus and nitrogen behaviour in the Mediterranean Sea. Deep Sea Research Part A. Oceanographic Research Papers, 39(9), pp.1641-1654.

Bible, A.N., Khalsa-Moyers, G.K., Mukherjee, T., Green, C.S., Mishra, P., Purcell, A., Aksenova, A., Hurst, G.B. and Alexandre, G., 2015. Metabolic adaptations of Azospirillum brasilense to oxygen stress by cell-to-cell clumping and flocculation. Applied and environmental microbiology, 81(24), pp.8346-8357.

Bird, C. and Wyman, M., 2013. Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea. FEMS microbiology ecology, 84(1), pp.189-200.

Bird, C., Martinez, J.M., O'Donnell, A.G. and Wyman, M., 2005. Spatial distribution and transcriptional activity of an uncultured clade of planktonic diazotrophic γ-Proteobacteria in the Arabian Sea. Applied and environmental microbiology, 71(4), pp.2079-2085.

Blais, M., Tremblay, J., Jungblut, A. D., Gagnon, J., Martin, J., Thaler, M., Lovejoy, C., 2012. Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. Global Biogeochemical Cycles, 26(3), GB3022.

Boetzer, M. and Pirovano, W., 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC bioinformatics, 15(1), p.211.

Bombar, D., Moisander, P.H., Dippner, J.W., Foster, R.A., Voss, M., Karfeld, B. and Zehr, J.P., 2011. Distribution of diazotrophic microorganisms and nifH gene expression in the Mekong River plume during intermonsoon. Marine Ecology Progress Series, 424, pp.39-52.

Bombar, D., Paerl, R.W. and Riemann, L., 2016. Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. Trends in microbiology, 24(11), pp.916-927.

Bombar, D., Turk-Kubo, K.A., Robidart, J., Carter, B.J. and Zehr, J.P., 2013. Non-cyanobacterial nifH phylotypes in the North Pacific Subtropical Gyre detected by flow-cytometry cell sorting. Environmental microbiology reports, 5(5), pp.705-715.

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K.A., Moutin, T., Hamersley, R.M., Grosso, O., Zehr, J.P. and Capone, D.G., 2013. Aphotic N 2 fixation in the eastern tropical South Pacific Ocean. PloS one, 8(12), p.e81265.

Bonnet, S., Grosso, O. and Moutin, T., 2011. Planktonic N_2 fixation in the Mediterranean Sea: a major biogeochemical process during the stratified period. Biogeosciences Discussions, 8(1), pp.1197-1225.

Bonnet, S., Guieu, C., Bruyant, F., Prášil, O., Van Wambeke, F., Raimbault, P., Moutin, T., Grob, C., Gorbunov, M.Y., Zehr, J.P. and Masquelier, S.M., 2008. Nutrient limitation of primary productivity in the Southeast Pacific (BIOSOPE cruise). Biogeosciences, 5(1), pp.215-225.

Bonnet, S., Guieu, C., Chiaverini, J., Ras, J., Stock, A., 2005. Effect of atmospheric nutrients on the autotrophic communities in a low nutrient, low chlorophyll system. Limnology and Oceanography, 50(6), pp.1810-1819.

Boström, K.H., Riemann, L., Kühl, M. and Hagström, Å., 2007. Isolation and gene quantification of heterotrophic N_2-fixing bacterioplankton in the Baltic Sea. Environmental microbiology, 9(1), pp.152-164.

Bowman, J.S., Rasmussen, S., Blom, N., Deming, J.W., Rysgaard, S. and Sicheritz-Ponten, T., 2012. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. The ISME journal, 6(1), pp.11-20.

Breitbarth, E., Oschlies, A., LaRoche, J., 2007. Physiological constraints on the global distribution of *Trichodesmium*–effect of temperature on diazotrophy. Biogeosciences, 4(1), pp.53-61.

Buchan, A., González, J.M. and Moran, M.A., 2005. Overview of the marine Roseobacter lineage. Applied and environmental microbiology, 71(10), pp.5665-5677.

Buck, C. S., Landing, W. M., Resing, J. A., Measures, C. I., 2010. The solubility and deposition of aerosol Fe and other trace elements in the North Atlantic Ocean: Observations from the A16N CLIVAR/CO2 repeat hydrography section. Marine Chemistry, 120(1), pp.57-70.

Bunse, C., Lundin, D., Karlsson, C.M., Akram, N., Vila-Costa, M., Palovaara, J., Svensson, L., Holmfeldt, K., González, J.M., Calvo, E. and Pelejero, C., 2016. Response of marine bacterioplankton pH homeostasis gene expression to elevated CO2. Nature Climate Change, 6(5), pp.483-487.

Burgess, B.K. and Lowe, D.J., 1996. Mechanism of molybdenum nitrogenase. Chemical Reviews, 96(7), pp.2983-3012.

Cabiscol, E., Tamarit, J. and Ros, J., 2010. Oxidative stress in bacteria and protein damage by reactive oxygen species. International Microbiology, 3(1), pp.3-8..

Cáceres, M.D. and Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology, 90(12), pp.3566-3574.

Calbet, A. and Landry, M.R., 2004. Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. Limnology and Oceanography, 49(1), pp.51-57.

Capone, D.G., Burns, J.A., Montoya, J.P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A.F. and Carpenter, E.J., 2005. Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. Global Biogeochemical Cycles, 19(2).

Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B. and Carpenter, E.J., 1997. *Trichodesmium*, a globally significant marine cyanobacterium. Science, 276(5316), pp.1221-1229.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R., 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics, 26(2), pp.266-267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I. and Huttley, G.A., 2010b. QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7(5), pp.335-336.

Carpenter, E. J., Montoya, J. P., Burns, J., Mulholland, M. R., Subramaniam, A., Capone, D. G., 1999. Extensive bloom of a N_2-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. Marine Ecology Progress Series, 185, pp.273-283.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), pp.540-552.

Chen, Y.B., Zehr, J.P. and Mellon, M., 1996. Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* spp. IMS 101 in defined media: evidence for a circadian rhythm. Journal of Phycology, 32(6), pp.916-923.

Cheung, S., Xia, X., Guo, C. and Liu, H., 2016. Diazotroph community structure in the deep oxygen minimum zone of the Costa Rica Dome. Journal of plankton research, 38(2), pp.380-391.

Cho, J.C. and Giovannoni, S.J., 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria. Applied and Environmental Microbiology, 70(1), pp.432-440.

Chow, C.E.T., Sachdeva, R., Cram, J.A., Steele, J.A., Needham, D.M., Patel, A., Parada, A.E. and Fuhrman, J.A., 2013. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. The ISME journal, 7(12), pp.2259-2273.

Church, M. J., Bjorkman, K. M., Karl, D. M., Saito, M. A., Zehr, J. P., 2008. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. Limnology and Oceanography, 53(1), p.63.

Church, M. J., Jenkins, B. D., Karl, D. M., Zehr, J. P., 2005. Vertical distributions of nitrogen-fixing phylotypes at stn ALOHA in the oligotrophic North Pacific Ocean. Aquatic Microbial Ecology, 38(1), pp.3-14.

Church, M.J., Mahaffey, C., Letelier, R.M., Lukas, R., Zehr, J.P. and Karl, D.M., 2009. Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. Global Biogeochemical Cycles, 23(2).

Clarke K. R., Gorley R. N., 2006. PRIMER v6: User manual/tutorial. PRIMER-E, Plymouth.

Codispoti, L., 2006. An oceanic fixed nitrogen sink exceeding 400 tg N a−1 vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences Discussions, 3(4), pp.1203-1246.

Codispoti, L.A., 2007. An oceanic fixed nitrogen sink exceeding 400 Tg N a-1 vs the concept of homeostasis in the fixed-nitrogen inventory. Biogeosciences, 4(2), pp.233-253.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W. and Yoshinari, T., 2001. The oceanic fixed nitrogen and nitrous

oxide budgets: Moving targets as we enter the anthropocene?. Scientia Marina, 65(S2), pp.85-105.

Comeau, A.M., Douglas, G.M. and Langille, M.G., 2017. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. mSystems, 2(1), pp.e00127-16.

Comeau, A.M., Li, W.K., Tremblay, J.É., Carmack, E.C. and Lovejoy, C., 2011. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLoS One, 6(11), p.e27492.

Conley, D.J., Humborg, C., Rahm, L., Savchuk, O.P. and Wulff, F., 2002. Hypoxia in the Baltic Sea and basin-scale changes in phosphorus biogeochemistry. Environmental science & technology, 36(24), pp.5315-5320.

Conway, T. M., John, S. G., 2014. Quantification of dissolved iron sources to the North Atlantic Ocean. Nature 511, pp.212–215.

Cook, P.L., Evrard, V. and Woodland, R.J., 2015. Factors controlling nitrogen fixation in temperate seagrass beds. Marine Ecology Progress Series, 525, pp.41-51.

Cram, J.A., Chow, C.E.T., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A. and Fuhrman, J.A., 2015b. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. The ISME journal, 9(3), pp.563-580.

Cram, J.A., Xia, L.C., Needham, D.M., Sachdeva, R., Sun, F. and Fuhrman, J.A., 2015a. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. The ISME journal, 9(12), pp.2573-2586.

Dahl, C., Engels, S., Pott-Sperling, A.S., Schulte, A., Sander, J., Lübbe, Y., Deuster, O. and Brune, D.C., 2005. Novel genes of the dsr gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium Allochromatium vinosum. Journal of bacteriology, 187(4), pp.1392-1404.

Dammshäuser, A., Wagener, T., Croot, P. L., 2011. Surface water dissolved aluminum and titanium: Tracers for specific time scales of dust deposition to the Atlantic. Geophysical Research Letters, 38(24) L24601.

Daneri, G., Dellarossa, V., Quiñones, R., Jacob, B., Montero, P. and Ulloa, O., 2000. Primary production and community respiration in the Humboldt Current System off Chile and associated oceanic areas. Marine Ecology Progress Series, 197, pp.41-49.

Davis, C.S. and McGillicuddy, D.J., 2006. Transatlantic abundance of the N2-fixing colonial cyanobacterium *Trichodesmium*. Science, 312(5779), pp.1517-1520.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I. and Carmichael, M., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science, 348(6237), p.1261605.

Decelle, J., Romac, S., Stern, R.F., Bendif, E.M., Zingone, A., Audic, S., Guiry, M.D., Guillou, L., Tessier, D., Le Gall, F. and Gourvil, P., 2015. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic

eukaryotes with curated taxonomy. Molecular ecology resources, 15(6), pp.1435-1445.

Dekaezemacker, J., Bonnet, S., Grosso, O., Moutin, T., Bressac, M. and Capone, D.G., 2013. Evidence of active dinitrogen fixation in surface waters of the eastern tropical South Pacific during El Niño and La Niña events and evaluation of its potential nutrient controls. Global Biogeochemical Cycles, 27(3), pp.768-779.

Dekas, A.E., Poretsky, R.S. and Orphan, V.J., 2009. Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. Science, 326(5951), pp.422-426.

Desai, D.K., Desai, F.D. and LaRoche, J., 2012. Factors influencing the diversity of iron uptake systems in aquatic microorganisms. Environmental Bioinorganic Chemistry of Aquatic Microbial Organisms, p.103.

Desai, M.S. and Brune, A., 2012. Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. The ISME journal, 6(7), pp.1302-1313.

Desai, D.K., Schunck, H., Löser, J.W. and LaRoche, J., 2013. Fragment recruitment on metabolic pathways: comparative metabolic profiling of metagenomes and metatranscriptomes. Bioinformatics, 29(6), pp.790-791.

Deutsch, C., Sarmiento, J.L., Sigman, D.M., Gruber, N. and Dunne, J.P., 2007. Spatial coupling of nitrogen inputs and losses in the ocean. Nature, 445(7124), pp.163-167.

Diaz, R.J. and Rosenberg, R., 2008. Spreading dead zones and consequences for marine ecosystems. science, 321(5891), pp.926-929.

Díez, B., Bergman, B., Pedrós-Alió, C., Antó, M. and Snoeijs, P., 2012. High cyanobacterial nifH gene diversity in Arctic seawater and sea ice brine. Environmental microbiology reports, 4(3), pp.360-366.

Dingler, C. and Oelze, J., 1987. Superoxide dismutase and catalase in Azotobacter vinelandii grown in continuous culture at different dissolved oxygen concentrations. Archives of microbiology, 147(3), pp.291-294.

Dingler, C., Kuhla, J., Wassink, H. and Oelze, J., 1988. Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations. Journal of bacteriology, 170(5), pp.2148-2152.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14(6), pp.927-930.

Dixon, R. and Kahn, D., 2004. Genetic regulation of biological nitrogen fixation. Nature Reviews Microbiology, 2(8), pp.621-631.

Donaghay, P. L., Liss, P. S., Duce, R. A., Kester, D. R., Hanson, A. K., Villareal, T., Gifford, D. J., 1991. The role of episodic atmospheric nutrient inputs in the chemical and biological dynamics of oceanic ecosystems. Oceanography, 4(2), pp.62-70.

Doney, S.C., Ruckelshaus, M., Duffy, J.E., Barry, J.P., Chan, F., English, C.A., Galindo, H.M., Grebmeier, J.M., Hollowed, A.B., Knowlton, N. and Polovina, J., 2012. Climate change impacts on marine ecosystems. Marine Science, 4.

Dong, S., Yang, J., Zhang, X.Y., Shi, M., Song, X.Y., Chen, X.L. and Zhang, Y.Z., 2012. Cultivable alginate lyase-excreting bacteria associated with the arctic brown alga Laminaria. Marine drugs, 10(11), pp.2481-2491.

Doudna, J.A. and Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213), p.1258096.

Du, Z.J., Lv, G.Q., Rooney, A.P., Miao, T.T., Xu, Q.Q. and Chen, G.J., 2011. Agarivorans gilvus sp. nov. isolated from seaweed. International journal of systematic and evolutionary microbiology, 61(3), pp.493-496.

Duarte, C. M., Dachs, J., Llabrés, M., Alonso-Laita, P., Gasol, J. M., Tovar-Sánchez, A., Agustí, S., 2006. Aerosol inputs enhance new production in the subtropical northeast atlantic. Journal of Geophysical Research, 111(G4), G04006.

Dubinsky, E.A., Conrad, M.E., Chakraborty, R., Bill, M., Borglin, S.E., Hollibaugh, J.T., Mason, O.U., M. Piceno, Y., Reid, F.C., Stringfellow, W.T. and Tom, L.M., 2013. Succession of hydrocarbon-degrading bacteria in the aftermath of the Deepwater Horizon oil spill in the Gulf of Mexico. Environmental science & technology, 47(19), pp.10860-10867.

Duce, R. A., LaRoche, J., Altieri, K., Arrigo, K. R., Baker, A. R., Capone, D. G., Zamora, L., 2008. Impacts of atmospheric anthropogenic nitrogen on the open ocean. Science, 320(5878), pp.893-897.

Dyhrman, S.T., Chappell, P.D., Haley, S.T., Moffett, J.W., Orchard, E.D., Waterbury, J.B. and Webb, E.A., 2006. Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. Nature, 439(7072), pp.68-71.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), pp.1792-1797.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27(16), pp.2194-2200.

Eisenhut, M., Kahlon, S., Hasse, D., Ewald, R., Lieman-Hurwitz, J., Ogawa, T., Ruth, W., Bauwe, H., Kaplan, A. and Hagemann, M., 2006. The plant-like C2 glycolate cycle and the bacterial-like glycerate pathway cooperate in phosphoglycolate metabolism in cyanobacteria. Plant Physiology, 142(1), pp.333-342.

El-Swais, H., Dunn, K.A., Bielawski, J.P., Li, W.K. and Walsh, D.A., 2015. Seasonal assemblages and short-lived blooms in coastal north-west Atlantic Ocean bacterioplankton. Environmental microbiology, 17(10), pp.3642-3661.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H. and Hugenholtz, P., 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. The ISME journal, 4(5), pp.642-647.

Entner, N. and Doudoroff, M., 1952. Glucose and gluconic acid oxidation of Pseudomonas saccharophila. Journal of Biological Chemistry, 196(2), pp.853-862.

Falcón, L.I., Carpenter, E.J., Cipriano, F., Bergman, B. and Capone, D.G., 2004. N_2 fixation by unicellular bacterioplankton from the Atlantic and Pacific Oceans:

Phylogeny and in situ rates. Applied and Environmental Microbiology, 70(2), pp.765-770.

Falkowski, P. G., 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. Nature, 387(6630), pp.272-275.

Falkowski, P.G., 1983. Enzymology of nitrogen assimilation. Nitrogen in the marine environment, pp.839-868.

Falkowski, P.G., Fenchel, T. and Delong, E.F., 2008. The microbial engines that drive Earth's biogeochemical cycles. science, 320(5879), pp.1034-1039.

Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å. and Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. PLoS One, 6(4), p.e19223.

Farnelid, H., Bentzon-Tilia, M., Andersson, A.F., Bertilsson, S., Jost, G., Labrenz, M., Jürgens, K. and Riemann, L., 2013. Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. The ISME journal, 7(7), pp.1413-1423.

Farnelid, H., Harder, J., Bentzon-Tilia, M. and Riemann, L., 2014. Isolation of heterotrophic diazotrophic bacteria from estuarine surface waters. Environmental microbiology, 16(10), pp.3072-3082.

Farnelid, H., Öberg, T. and Riemann, L., 2009. Identity and dynamics of putative N_2-fixing picoplankton in the Baltic Sea proper suggest complex patterns of regulation. Environmental microbiology reports, 1(2), pp.145-154.

Farnelid, H., Tarangkoon, W., Hansen, G., Hansen, P.J. and Riemann, L., 2010. Putative N_2-fixing heterotrophic bacteria associated with dinoflagellate–Cyanobacteria consortia in the low-nitrogen Indian Ocean. Aquatic Microbial Ecology, 61(2), pp.105-117.

Favinger, J., Stadtwald, R. and Gest, H., 1989. Rhodospirillum centenum, sp. nov., a thermotolerant cyst-forming anoxygenic photosynthetic bacterium. Antonie van Leeuwenhoek, 55(3), pp.291-296.

Fernández, A., Graña, R., Mouriño-Carballido, B., Bode, A., Varela, M., Domínguez-Yanes, J. F., Marañón, E., 2013. Community N_2 fixation and *Trichodesmium* spp. abundance along longitudinal gradients in the eastern subtropical North Atlantic. ICES Journal of Marine Science: Journal Du Conseil, 70(1), pp.223-231.

Fernandez, C., Farías, L. and Ulloa, O., 2011. Nitrogen fixation in denitrified marine waters. PLoS One, 6(6), p.e20539.

Fernandez, C., González, M.L., Muñoz, C., Molina, V. and Farias, L., 2015. Temporal and spatial variability of biological nitrogen fixation off the upwelling system of central Chile (35–38.5° S). Journal of Geophysical Research: Oceans, 120(5), pp.3330-3349.

Fernández-Méndez, M., Turk-Kubo, K.A., Buttigieg, P.L., Rapp, J.Z., Krumpen, T., Zehr, J.P. and Boetius, A., 2016. Diazotroph Diversity in the Sea Ice, Melt Ponds, and Surface Waters of the Eurasian Basin of the Central Arctic Ocean. Frontiers in Microbiology, 7.

Fitzsimmons, J. N., Zhang, R., Boyle, E. A., 2013. Dissolved iron in the tropical north atlantic ocean. Marine Chemistry, 154, pp.87-99.

Forth, M., Liljebladh, B., Stigebrandt, A., Hall, P.O. and Treusch, A.H., 2015. Effects of ecological engineered oxygenation on the bacterial community structure in an anoxic fjord in western Sweden. The ISME journal, 9(3), pp.656-669.

Foster, R., Subramaniam, A., Mahaffey, C., Carpenter, E., Capone, D., Zehr, J., 2007. Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical North Atlantic Ocean. Limnology and Oceanography, 52(2), pp.517-532.

Foster, R.A., Paytan, A. and Zehr, J.P., 2009. Seasonality of N_2 fixation and nifH gene diversity in the Gulf of Aqaba (Red Sea). Limnology and Oceanography, 54(1), pp.219-233.

Fouilland, E., Tolosa, I., Bonnet, D., Bouvier, C., Bouvier, T., Bouvy, M., Got, P., Le Floc'h, E., Mostajir, B., Roques, C. and Sempéré, R., 2014. Bacterial carbon dependence on freshly produced phytoplankton exudates under different nutrient availability and grazing pressure conditions in coastal marine waters. FEMS microbiology ecology, 87(3), pp.757-769.

Franchy, G., Ojeda, A., López-Cancio, J., Hernández-León, S., 2013. Plankton community response to Saharan dust fertilization in subtropical waters off the Canary Islands. Biogeosciences Discussions, 10(11), pp.17275-17307.

Frankel, R.B., Bazylinski, D.A., Johnson, M.S. and Taylor, B.L., 1997. Magneto-aerotaxis in marine coccoid bacteria. Biophysical journal, 73(2), p.994.

Fredriksson, C. and Bergman, B., 1997. Ultrastructural characterisation of cells specialised for nitrogen fixation in a non-heterocystous cyanobacterium, *Trichodesmium* spp. Protoplasma, 197(1-2), pp.76-85.

Friedman, J. and Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol, 8(9), p.e1002687.

Fu, Y., Keats, K.F., Rivkin, R.B. and Lang, A.S., 2013. Water mass and depth determine the distribution and diversity of Rhodobacterales in an Arctic marine system. FEMS microbiology ecology, 84(3), pp.564-576.

Fuhrman, J.A., Cram, J.A. and Needham, D.M., 2015. Marine microbial community dynamics and their ecological interpretation. Nature Reviews Microbiology, 13(3), pp.133-146.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V. and Naeem, S., 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences, 103(35), pp.13104-13109.

Fuller, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F. and Scanlan, D.J., 2005. Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. Limnology and Oceanography, 50(1), pp.363-375.

Gamble, M.D., Bagwell, C.E., LaRocque, J., Bergholz, P.W. and Lovell, C.R., 2010. Seasonal variability of diazotroph assemblages associated with the

rhizosphere of the salt marsh cordgrass, Spartina alterniflora. Microbial ecology, 59(2), pp.253-265.

Gao, Y., Kaufman, Y., Tanre, D., Kolber, D., Falkowski, P., 2001. Seasonal distributions of aeolian iron fluxes to the global ocean. Geophysical Research Letters, 28(1), pp.29-32.

Gauthier, M.J. and Flatau, G.N., 1976. Antibacterial activity of marine violet-pigmented Alteromonas with special reference to the production of brominated compounds. Canadian Journal of Microbiology, 22(11), pp.1612-1619.

Gauthier, M.J., 1977. Alteromonas citrea, a new Gram-negative, yellow-pigmented species from seawater. International Journal of Systematic and Evolutionary Microbiology, 27(4), pp.349-354.

Georges, A.A., El-Swais, H., Craig, S.E., Li, W.K. and Walsh, D.A., 2014. Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. ISME J, 8(6), pp.1301-1313.

Ghiglione, J.F., Galand, P.E., Pommier, T., Pedrós-Alió, C., Maas, E.W., Bakker, K., Bertilson, S., Kirchman, D.L., Lovejoy, C., Yager, P.L. and Murray, A.E., 2012. Pole-to-pole biogeography of surface and deep marine bacterial communities. Proceedings of the National Academy of Sciences, 109(43), pp.17633-17638.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I. and Somerfield, P., 2012. Defining seasonal marine microbial community dynamics. The ISME journal, 6(2), pp.298-308.

Giovannoni, S.J. and Vergin, K.L., 2012. Seasonality in ocean microbial communities. Science, 335(6069), pp.671-676.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M. and Rappé, M.S., 2005. Genome streamlining in a cosmopolitan oceanic bacterium. science, 309(5738), pp.1242-1245.

Giuffrè, A., Borisov, V.B., Arese, M., Sarti, P. and Forte, E., 2014. Cytochrome bd oxidase and bacterial tolerance to oxidative and nitrosative stress. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1837(7), pp.1178-1187.

Glaubitz, S., Labrenz, M., Jost, G. and Jürgens, K., 2010. Diversity of active chemolithoautotrophic prokaryotes in the sulfidic zone of a Black Sea pelagic redoxcline as determined by rRNA-based stable isotope probing. FEMS microbiology ecology, 74(1), pp.32-41.

Glaubitz, S., Lueders, T., Abraham, W.R., Jost, G., Jürgens, K. and Labrenz, M., 2009. 13C-isotope analyses reveal that chemolithoautotrophic Gamma-and Epsilonproteobacteria feed a microbial food web in a pelagic redoxcline of the central Baltic Sea. Environmental microbiology, 11(2), pp.326-337.

Goebel, N. L., Edwards, C. A., Church, M. J., Zehr, J. P., 2007. Modeled contributions of three types of diazotrophs to nitrogen fixation at station ALOHA. The ISME Journal, 1(7), pp.606-619.

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya, J. P., Edwards, C. A., Zehr, J. P., 2010. Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N_2 fixation in the tropical Atlantic Ocean. Environmental Microbiology, 12(12), pp.3272-3272.

Golyshin, P.N., Werner, J., Chernikova, T.N., Tran, H., Ferrer, M., Yakimov, M.M., Teeling, H., Golyshina, O.V. and MAMBA Scientific Consortium, 2013. Genome sequence of Thalassolituus oleivorans MIL-1 (DSM 14913T). Genome announcements, 1(2), pp.e00141-13.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M.C. and McCombie, W.R., 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome research, 25(11), pp.1750-1756.

Gosink, J.J., Woese, C.R. and Staley, J.T., 1998. Polaribacter gen. nov., with three new species, P. irgensii sp. nov., P. franzmannii sp. nov. and P. filamentus sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of 'Flectobacillus glomeratus' as Polaribacter glomeratus comb. nov. International Journal of Systematic and Evolutionary Microbiology, 48(1), pp.223-235.

Goto, M., Ando, S., Hachisuka, Y. and Yoneyama, T., 2005. Contamination of diverse nifH and nifH-like DNA into commercial PCR primers. FEMS microbiology letters, 246(1), pp.33-38.

Goudie, A., Middleton, N., 2001. Saharan dust storms: Nature and consequences. Earth-Science Reviews, 56(1), pp.179-204.

Gram, L., Melchiorsen, J. and Bruhn, J.B., 2010. Antibacterial activity of marine culturable bacteria collected from a global sampling of ocean surface waters and surface swabs of marine organisms. Marine biotechnology, 12(4), pp.439-451.

Grand, M.M., Buck, C. S., Landing W.M., Measures, C.I., Hatta, M., Hiscock, W.T., Brown, M., Resing. J.A., 2014. Quantifying the impact of atmospheric deposition on the biogeochemistry of Fe and Al in the upper ocean: A decade of collaboration with the US CLIVAR-CO2 Repeat Hydrography Program. Oceanography 27(1): pp.62–65.

Gregory, M.R., 2009. Environmental implications of plastic debris in marine settings—entanglement, ingestion, smothering, hangers-on, hitch-hiking and alien invasions. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364(1526), pp.2013-2025.

Grice, E.A. and Segre, J.A., 2011. The skin microbiome. Nature Reviews Microbiology, 9(4), pp.244-253.

Großkopf, T. and LaRoche, J., 2012. Direct and indirect costs of N_2 fixation in Crocosphaera watsonii WH8501 and possible implications for the nitrogen cycle. Frontiers in microbiology, 3, p.236.

Großkopf, T., Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M.M., Lavik, G., Schmitz, R.A., Wallace, D.W. and LaRoche, J., 2012. Doubling of marine N_2-fixation rates based on direct measurements. Nature, 488(7411), pp.361-364.

Gruber, N., 2008. The marine nitrogen cycle: overview and challenges. Nitrogen in the marine environment, pp.1-50.

Gruber, N., Sarmiento, J. L., 1997. Global patterns of marine nitrogen fixation and denitrification. Global Biogeochemical Cycles, 11(2), pp.235-266.

Guerzoni, S., Chester, R., Dulac, F., Herut, B., Loÿe-Pilot, M., Migon, C., Saydam, C., 1999. The role of atmospheric deposition in the biogeochemistry of the Mediterranean Sea. Progress in Oceanography, 44(1), pp.147-190.

Guillard, R.R. and Ryther, J.H., 1962. Studies of marine planktonic diatoms: I. Cyclotella Nana Hustedt, and Detonula Confervacea (CLEVE) Gran. Canadian journal of microbiology, 8(2), pp.229-239.

Gutierrez, T., Singleton, D.R., Berry, D., Yang, T., Aitken, M.D. and Teske, A., 2013. Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. The ISME journal, 7(11), pp.2091-2104.

Halm, H., Lam, P., Ferdelman, T.G., Lavik, G., Dittmar, T., LaRoche, J., D'Hondt, S. and Kuypers, M.M., 2012. Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. The ISME journal, 6(6), pp.1238-1249.

Hamersley, M.R., Lavik, G., Woebken, D., Rattray, J.E., Lam, P., Hopmans, E.C., Damsté, J.S.S., Krüger, S., Graco, M., Gutiérrez, D. and Kuypers, M.M., 2007. Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. Limnology and Oceanography, 52(3), pp.923-933.

Hamersley, M.R., Turk, K.A., Leinweber, A., Gruber, N., Zehr, J.P., Gunderson, T. and Capone, D.G., 2011. Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. Aquatic Microbial Ecology, 63(2), pp.193-205.

Haselkorn, R., 2007. Heterocyst differentiation and nitrogen fixation in cyanobacteria. In Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations (pp. 233-255). Springer Netherlands.

Hashimoto, R., Yoshida, T., Kuno, S., Nishikawa, T. and Sako, Y., 2012. The first assessment of cyanobacterial and diazotrophic diversities in the Japan Sea. Fisheries Science, 78(6), pp.1293-1300.

Hellou, J., Steller, S., Zitko, V., Leonard, J., King, T., Milligan, T.G. and Yeats, P., 2002. Distribution of PACs in surficial sediments and bioavailability to mussels, Mytilus edulis of Halifax Harbour. Marine environmental research, 53(4), pp.357-379.

Herut, B., Zohary, T., Krom, M., Mantoura, R. F. C., Pitta, P., Psarra, S., Frede Thingstad, T., 2005. Response of east Mediterranean surface water to Saharan dust: On-board microcosm experiment and field observations. Deep Sea Research Part II: Topical Studies in Oceanography, 52(22), pp.3024-3040.

Hewson, I., Moisander, P.H., Achilles, K.M., Carlson, C.A., Jenkins, B.D., Mondragon, E.A., Morrison, A.E. and Zehr, J.P., 2007. Characteristics of diazotrophs in surface to abyssopelagic waters of the Sargasso Sea. Aquatic microbial ecology, 46(1), pp.15-30.

Hilton, J.A., Satinsky, B.M., Doherty, M., Zielinski, B. and Zehr, J.P., 2015. Metatranscriptomics of N_2-fixing cyanobacteria in the Amazon River plume. The ISME journal, 9(7), pp.1557-1569.

Hoeft McCann, S., Boren, A., Hernandez-Maldonado, J., Stoneburner, B., Saltikov, C.W., Stolz, J.F. and Oremland, R.S., 2016. Arsenite as an Electron Donor for Anoxygenic Photosynthesis: Description of Three Strains of Ectothiorhodospira from Mono Lake, California and Big Soda Lake, Nevada. Life, 7(1), p.1.

Hoffman, B.M., Lukoyanov, D., Yang, Z.Y., Dean, D.R. and Seefeldt, L.C., 2014. Mechanism of nitrogen fixation by nitrogenase: the next stage. Chemical reviews, 114(8), pp.4041-4062.

Holl, C.M., Waite, A.M., Pesant, S., Thompson, P.A. and Montoya, J.P., 2007. Unicellular diazotrophy as a source of nitrogen to Leeuwin Current coastal eddies. Deep Sea Research Part II: Topical Studies in Oceanography, 54(8), pp.1045-1054.

Huete-Stauffer, T.M., Arandia-Gorostidi, N., Díaz-Pérez, L. and Moran, X.A.G., 2015. Temperature dependences of growth rates and carrying capacities of marine bacteria depart from metabolic theoretical predictions. FEMS microbiology ecology, 91(10), p.fiv111.

Huston, A.L., 2003. Bacterial adaptation to the cold: in Sim aotivities of extracellular (Doctoral dissertation, University of Washington).

Huston, A.L., Krieger-Brockett, B.B. and Deming, J.W., 2000. Remarkably low temperature optima for extracellular enzyme activity from Arctic bacteria and sea ice. Environmental Microbiology, 2(4), pp.383-388.

Huston, A.L., Methe, B. and Deming, J.W., 2004. Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile Colwellia psychrerythraea strain 34H. Applied and environmental microbiology, 70(6), pp.3321-3328.

Hutchins, D.A., Mulholland, M.R. and Fu, F., 2009. Nutrient cycles and marine microbes in a CO2-enriched ocean. Oceanography.

Imai, I., Ishida, Y., Sakaguchi, K. and Hata, Y., 1995. Algicidal marine bacteria isolated from northern Hiroshima Bay, Japan. Fisheries science, 61(4), pp.628-636.

Imlay, J.A., Chin, S.M. and Linn, S., 1988. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. Science, 240(4852), p.640.

Ingall, E. and Jahnke, R., 1994. Evidence for enhanced phosphorus regeneration from marine sediments overlain by oxygen depleted waters. Geochimica et Cosmochimica Acta, 58(11), pp.2571-2575.

Inomura, K., Bragg, J. and Follows, M.J., 2017. A quantitative analysis of the direct and indirect costs of nitrogen fixation: a model based on Azotobacter vinelandii. The ISME Journal, 11(1), pp.166-175.

Izquierdo, J.A. and Nüsslein, K., 2006. Distribution of extensive nifH gene diversity across physical soil microenvironments. Microbial Ecology, 51(4), pp.441-452.

Jacoby, W.G., 2000. Loess:: a nonparametric, graphical tool for depicting relationships between variables. Electoral Studies, 19(4), pp.577-613.

Jahson, S., Rai, A.N. and Bergman, B., 1995. Intracellular cyanobiont Richelia intracellularis: ultrastructure and immuno-localisation of phycoerythrin, nitrogenase, Rubisco and glutamine synthetase. Marine Biology, 124(1), pp.1-8.

Janssen, F., Schrum, C. and Backhaus, J.O., 1999. A climatological data set of temperature and salinity for the Baltic Sea and the North Sea. Deutsche Hydrografische Zeitschrift, 51(9), p.5.

Jayakumar, A., Al-Rshaidat, M.M., Ward, B.B. and Mulholland, M.R., 2012. Diversity, distribution, and expression of diazotroph nifH genes in oxygen-deficient waters of the Arabian Sea. FEMS microbiology ecology, 82(3), pp.597-606.

Jickells, T., 1999. The inputs of dust derived elements to the Sargasso Sea; a synthesis. Marine Chemistry, 68(1), pp.5-14.

Jickells, T.D., An, Z.S., Andersen, K.K., Baker, A.R., Bergametti, G., Brooks, N., Cao, J.J., Boyd, P.W., Duce, R.A., Hunter, K.A. and Kawahata, H., 2005. Global iron connections between desert dust, ocean biogeochemistry, and climate. science, 308(5718), pp.67-71.

Johnston, A.W., Li, Y. and Ogilvie, L., 2005. Metagenomic marine nitrogen fixation–feast or famine?. Trends in microbiology, 13(9), pp.416-420.

Joint, I., Tait, K., Callow, M.E., Callow, J.A., Milton, D., Williams, P. and Cámara, M., 2002. Cell-to-cell communication across the prokaryote-eukaryote boundary. Science, 298(5596), pp.1207-1207.

Juenemann, S., Butterworth, P.J. and Wrigglesworth, J.M., 1995. A suggested mechanism for the catalytic cycle of cytochrome bd terminal oxidase based on kinetic analysis. Biochemistry, 34(45), pp.14861-14867.

Jungblut, A.D., Lovejoy, C. and Vincent, W.F., 2010. Global distribution of cyanobacterial ecotypes in the cold biosphere. The ISME Journal, 4(2), pp.191-202.

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D. and Stal, L., 2002. Dinitrogen fixation in the world's oceans. In The Nitrogen Cycle at Regional to Global Scales (pp. 47-98). Springer Netherlands.

Karl, D.M. and Church, M.J., 2014. Microbial oceanography and the Hawaii Ocean Time-series programme. Nature Reviews Microbiology, 12(10), pp.699-713.

Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research, 30(14), pp.3059-3066.

Kim, H., Choo, Y.J., Song, J., Lee, J.S., Lee, K.C. and Cho, J.C., 2007. Marinobacterium litorale sp. nov. in the order Oceanospirillales. International journal of systematic and evolutionary microbiology, 57(7), pp.1659-1662.

Kirchman, D.L., 2000. Uptake and regeneration of inorganic nutrients by marine heterotrophic bacteria. Microbial ecology of the oceans, pp.261-288.

Klindworth, A., Mann, A.J., Huang, S., Wichels, A., Quast, C., Waldmann, J., Teeling, H. and Glöckner, F.O., 2014. Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. Marine genomics, 18, pp.185-192.

Knapp, A.N., 2015. The sensitivity of marine N 2 fixation to dissolved inorganic nitrogen. The microbial nitrogen cycle, p.90.

Kolonay, J.F., Moshiri, F., Gennis, R.B., Kaysser, T.M. and Maier, R.J., 1994. Purification and characterization of the cytochrome bd complex from Azotobacter vinelandii: comparison to the complex from Escherichia coli. Journal of bacteriology, 176(13), pp.4177-4181.

Kong, L., Jing, H., Kataoka, T., Sun, J. and Liu, H., 2011. Phylogenetic diversity and spatio-temporal distribution of nitrogenase genes (nifH) in the northern South China Sea. Aquatic Microbial Ecology, 65(1), pp.15-27.

Kopylova, E., Noé, L. and Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics, 28(24), pp.3211-3217.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv, p.071282.

Krom, M.D., Kress, N., Brenner, S. and Gordon, L.I., 1991. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. Limnology and Oceanography, 36(3), pp.424-432.

Krupke, A., Mohr, W., LaRoche, J., Fuchs, B.M., Amann, R.I. and Kuypers, M.M., 2015. The effect of nutrients on carbon and nitrogen fixation by the UCYN-A–haptophyte symbiosis. The ISME journal, 9(7), pp.1635-1647.

Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., Foster, R. A., 2013. In situ identification and N_2 and C fixation rates of uncultivated cyanobacteria populations. Systematic and Applied Microbiology, 36(4), pp.259–271.

Kujawinski, E.B., Longnecker, K., Barott, K.L., Weber, R.J.M. and Kido Soule, M.C., 2016. Microbial Community Structure Affects Marine Dissolved Organic Matter Composition. Front. Mar. Sci, 3, p.45.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution, p.msw054.

Kustka, A.B., Sanudo-Wilhelmy, S.A., Carpenter, E.J., Capone, D., Burns, J. and Sunda, W.G., 2003. Iron requirements for dinitrogen-and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations. Limnology and Oceanography, 48(5), pp.1869-1884.

Labiosa, R.G., Arrigo, K.R., Genin, A., Monismith, S.G. and van Dijken, G., 2003. The interplay between upwelling and deep convective mixing in determining the seasonal phytoplankton dynamics in the Gulf of Aqaba: Evidence from SeaWiFS and MODIS. Limnology and oceanography, 48(6), pp.2355-2368.

Langlois, R. J., Hümmer, D., LaRoche, J., 2008. Abundances and distributions of the dominant nifH phylotypes in the northern Atlantic Ocean. Applied and Environmental Microbiology, 74(6), pp.1922-1931.

Langlois, R. J., LaRoche, J., Raab, P. A., 2005. Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. Applied and Environmental Microbiology, 71(12), pp.7910-7919.

Langlois, R. J., Mills, M. M., Ridame, C., Croot, P., LaRoche, J., 2012. Diazotrophic bacteria respond to Saharan dust additions. Marine Ecology Progress Series, pp.470, 1-14.

Langlois, R., Großkopf, T., Mills, M., Takeda, S. and LaRoche, J., 2015. Widespread distribution and expression of gamma A (UMB), an uncultured, diazotrophic, γ-proteobacterial nifH phylotype. PloS one, 10(6), p.e0128912.

Lawrence, C. R., Neff, J. C., 2009. The contemporary physical and chemical flux of aeolian dust: A synthesis of direct measurements of dust deposition. Chemical Geology, 267(1), pp.46-63.

Lee, T. N., Yoder, J. A., Atkinson, L. P., 1991. Gulf stream frontal eddy influence on productivity of the southeast US continental shelf. Journal of Geophysical Research: Oceans (1978–2012), 96(C12), pp.22191-22205.

Lema, K.A., Willis, B.L. and Bourne, D.G., 2012. Corals form characteristic associations with symbiotic nitrogen-fixing bacteria. Applied and environmental microbiology, 78(9), pp.3136-3144.

Letunic, I. and Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, p.gkw290.

Li, H., Lee, B.C., Kim, T.S., Bae, K.S., Hong, J.K., Choi, S.H., Bao, B. and Jung, J.H., 2008. Bioactive cyclic dipeptides from a marine sponge-associated bacterium, Psychrobacter sp. Biomolecules and Therapeutics, 16(4), pp.356-363.

Li, H., Shinde, P.B., Lee, H.J., Yoo, E.S., Lee, C.O., Hong, J., Choi, S.H. and Jung, J.H., 2009. Bile acid derivatives from a sponge-associated bacterium Psychrobacter sp. Archives of pharmacal research, 32(6), pp.857-862.

Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), pp.1658-1659.

Li, W. K., 2014. The state of phytoplankton and bacterioplankton at the Compass Buoy Station: Bedford Basin Monitoring Program 1992-2013. Canadian Technical Report of Hydrography and Ocean Sciences 304.

Li, W.K. and Glen Harrison, W., 2008. Propagation of an atmospheric climate signal to phytoplankton in a small marine basin. Limnology and Oceanography, 53(5), pp.1734-1745.

Li, W.K., Harrison, W.G. and Head, E.J., 2006. Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. Proceedings of the Royal Society of London B: Biological Sciences, 273(1596), pp.1953-1960.

Li, W.K.W. and Dickie, P.M., 2001. Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. Cytometry, 44(3), pp.236-246.

Loescher, C.R., Großkopf, T., Desai, F.D., Gill, D., Schunck, H., Croot, P.L., Schlosser, C., Neulinger, S.C., Pinnow, N., Lavik, G. and Kuypers, M.M., 2014. Facets of diazotrophy in the oxygen minimum zone waters off Peru. The ISME journal, 8(11), pp.2180-2192.

Lorenz, M.C. and Fink, G.R., 2002. Life and death in a macrophage: role of the glyoxylate cycle in virulence. Eukaryotic cell, 1(5), pp.657-662.

Loveless, T.M., Saah, J.R. and Bishop, P.E., 1999. Isolation of nitrogen-fixing bacteria containing molybdenum-independent nitrogenases from natural environments. Applied and environmental microbiology, 65(9), pp.4223-4226.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. and Knight, R., 2012. Diversity, stability and resilience of the human gut microbiota. Nature, 489(7415), pp.220-230.

Lucas, J., Wichels, A., Teeling, H., Chafee, M., Scharfe, M. and Gerdts, G., 2015. Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. FEMS microbiology ecology, 91(9), p.fiv099.

Luo, Y.W., Doney, S.C., Anderson, L.A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K.H., Böttjer, D., Capone, D.G. and Carpenter, E.J., 2012. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. Earth System Science Data, 4(1), pp.47-73.

Lynch, M.D. and Neufeld, J.D., 2015. Ecology and exploration of the rare biosphere. Nature Reviews Microbiology, 13(4), pp.217-229.

Madigan, M.T., 1995. Microbiology of nitrogen fixation by anoxygenic photosynthetic bacteria. In Anoxygenic photosynthetic bacteria (pp. 915-928). Springer Netherlands.

Madoui, M.A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P. and Aury, J.M., 2015. Genome assembly using Nanopore-guided long and error-free DNA reads. BMC genomics, 16(1), p.327.

Mahaffey, C., Michaels, A.F. and Capone, D.G., 2005. The conundrum of marine N_2 fixation. American Journal of Science, 305(6-8), pp.546-595.

Mahowald, N., Jickells, T.D., Baker, A.R., Artaxo, P., Benitez-Nelson, C.R., Bergametti, G., Bond, T.C., Chen, Y., Cohen, D.D., Herut, B. and Kubilay, N., 2008. Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. Global Biogeochemical Cycles, 22(4).

Man-Aharonovich, D., Kress, N., Zeev, E.B., Berman-Frank, I. and Béjà, O., 2007. Molecular ecology of nifH genes and transcripts in the eastern Mediterranean Sea. Environmental microbiology, 9(9), pp.2354-2363.

Manasrah, R., Zibdah, M., Al-Ougaily, F., Yusuf, N. and Al-Najjar, T., 2007. Seasonal changes of water properties and current in the northernmost Gulf of Aqaba, Red Sea. Ocean Science Journal, 42(2), pp.103-116.

Maranón, E., Fernández, A., Mourino-Carballido, B., Martínez-García, S., Teira, E., Cermeno, P., Calvo-Díaz, A., 2010. Degree of oligotrophy controls the response of microbial plankton to Saharan dust. Limnology and Oceanography, 55(6), pp.2339-2352.

Marcello, J., Hernandez-Guerra, A., Eugenio, F., Fonte, A., 2011. Seasonal and temporal study of the northwest African upwelling system. International Journal of Remote Sensing, 32(7), pp.1843-1859.

Marchal, K. and Vanderleyden, J., 2000. The" oxygen paradox" of N_2-fixing bacteria. Biology and fertility of soils, 30(5-6), pp.363-373.

Martinez-Perez, C., Mohr, W., Löscher, C.R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B.M., Lavik, G., Schmitz, R.A. and LaRoche, J., 2016. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. Nature Microbiology, 1, p.16163.

Marty, D.G., 1993. Methanogenic bacteria in seawater. Limnology and oceanography, 38(2), pp.452-456.

Maruyama, Y., Taga, N. and Matsuda, O., 1970. Distribution of nitrogen-fixing bacteria in the central Pacific Ocean. Journal of the Oceanographical Society of Japan, 26(6), pp.360-366.

Marx, J.G., Carpenter, S.D. and Deming, J.W., 2009. Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium Colwellia psychrerythraea strain 34H under extreme conditions This article is one of a selection of papers in the Special Issue on Polar and Alpine Microbiology. Canadian journal of microbiology, 55(1), pp.63-72.

Mason, O.U., 2014. Metagenomics, metatranscriptomics and single cell genomics reveal functional response of active Oceanospirillales to Gulf oil spill. Journal of Microbial Ecology (The ISME).

Mason, O.U., Han, J., Woyke, T. and Jansson, J.K., 2015. Single-cell genomics reveals features of a Colwellia species that was dominant during the Deepwater Horizon oil spill. The metabolic pathways and environmental controls of hydrocarbon biodegradation in marine ecosystems, p.29.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K. and Pedrós-Alió, C., 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Applied and environmental microbiology, 70(6), pp.3528-3534.

McCarthy, S.A., Johnson, R.M. and Kakimoto, D., 1994. Characterization of an antibiotic produced by Alteromonas luteoviolacea Gauthier 1982, 85 isolated from Kinko Bay, Japan. Journal of applied bacteriology, 77(4), pp.426-432.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. and Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal, 6(3), pp.610-618.

McGrath, J.W., Chin, J.P. and Quinn, J.P., 2013. Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. Nature Reviews Microbiology, 11(6), pp.412-419.

McRose, D.L., Zhang, X., Kraepiel, A.M. and Morel, F.M., 2017. Diversity and Activity of Alternative Nitrogenases in Sequenced Genomes and Coastal Environments. Frontiers in Microbiology, 8, p.267.

Mehta, M.P. and Baross, J.A., 2006. Nitrogen fixation at 92 C by a hydrothermal vent archaeon. Science, 314(5806), pp.1783-1786.

Mehta, M.P., Butterfield, D.A. and Baross, J.A., 2003. Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge. Applied and Environmental Microbiology, 69(2), pp.960-970.

Mehta, M.P., Huber, J.A. and Baross, J.A., 2005. Incidence of novel and potentially archaeal nitrogenase genes in the deep Northeast Pacific Ocean. Environmental microbiology, 7(10), pp.1525-1534.

Mercier, C., Boyer, F., Bonin, A. and Coissac, E., 2013. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. Programs Abstr SeqBio, pp.27-9.

Mercurio, P., Flores, F., Mueller, J.F., Carter, S. and Negri, A.P., 2014. Glyphosate persistence in seawater. Marine pollution bulletin, 85(2), pp.385-390.

Methé, B.A., Nelson, K.E., Deming, J.W., Momen, B., Melamud, E., Zhang, X., Moult, J., Madupu, R., Nelson, W.C., Dodson, R.J. and Brinkac, L.M., 2005. The psychrophilic lifestyle as revealed by the genome sequence of Colwellia psychrerythraea 34H through genomic and proteomic analyses. Proceedings of the National Academy of Sciences of the United States of America, 102(31), pp.10913-10918.

Mills, M. M., Ridame, C., Davey, M., La Roche, J., Geider, R. J., 2004. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. Nature, 429(6989), pp.292-294.

Moeseneder, M.M., Arrieta, J.M. and Herndl, G.J., 2005. A comparison of DNA- and RNA-based clone libraries from the same marine bacterioplankton community. FEMS microbiology ecology, 51(3), pp.341-352.

Moffett, J.W., Goepfert, T.J. and Naqvi, S.W.A., 2007. Reduced iron associated with secondary nitrite maxima in the Arabian Sea. Deep Sea Research Part I: Oceanographic Research Papers, 54(8), pp.1341-1349.

Mohr, W., Intermaggio, M.P. and LaRoche, J., 2010. Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium Crocosphaera watsonii WH8501. Environmental microbiology, 12(2), pp.412-421.

Moisander, P.H., Beinart, R.A., Hewson, I., White, A.E., Johnson, K.S., Carlson, C.A., Montoya, J.P. and Zehr, J.P., 2010. Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. Science, 327(5972), pp.1512-1514.

Moisander, P.H., Beinart, R.A., Voss, M. and Zehr, J.P., 2008. Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. The ISME journal, 2(9), pp.954-967.

Moisander, P.H., Serros, T., Paerl, R.W., Beinart, R.A. and Zehr, J.P., 2014. Gammaproteobacterial diazotrophs and nifH gene expression in surface waters of the South Pacific Ocean. The ISME journal, 8(10), pp.1962-1973.

Monteiro, F.M., Follows, M.J. and Dutkiewicz, S., 2010. Distribution of diverse nitrogen fixers in the global ocean. Global Biogeochemical Cycles, 24(3).

Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., Rijkenberg, M. J., 2009. Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. Nature Geoscience, 2(12), pp.867-871.

Moore, C.M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., Galbraith, E.D., Geider, R.J., Guieu, C., Jaccard, S.L. and Jickells, T.D., 2013. Processes and patterns of oceanic nutrient limitation. Nature Geoscience, 6(9), pp.701-710.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A. and Giovannoni, S.J., 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature, 420(6917), pp.806-810.

Moshiri, F., Crouse, B.R., Johnson, M.K. and Maier, R.J., 1995. The" nitrogenase-protective" FeSII protein of Azotobacter vinelandii: overexpression, characterization, and crystallization. Biochemistry, 34(40), pp.12973-12982.

Moutin, T., Karl, D.M., Duhamel, S., Rimmelin, P., Raimbault, P., Van Mooy, B.A. and Claustre, H., 2007. Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean. Biogeosciences Discussions, 4(4), pp.2407-2440.

Mulholland, M.R., Bernhardt, P.W., Blanco-Garcia, J.L., Mannino, A., Hyde, K., Mondragon, E., Turk, K., Moisander, P.H. and Zehr, J.P., 2012. Rates of N_2 fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A. and Anson, E.L., 2000. A whole-genome assembly of Drosophila. Science, 287(5461), pp.2196-2204.

Naqvi, S.W.A., 2008. The Indian Ocean.

Nedashkovskaya, O.I., Kim, S.B., Han, S.K., Rhee, M.S., Lysenko, A.M., Falsen, E., Frolova, G.M., Mikhailov, V.V. and Bae, K.S., 2004. Ulvibacter litoralis gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the green alga Ulva fenestrata. International journal of systematic and evolutionary microbiology, 54(1), pp.119-123.

Nedashkovskaya, O.I., Kim, S.B., Vancanneyt, M., Snauwaert, C., Lysenko, A.M., Rohde, M., Frolova, G.M., Zhukova, N.V., Mikhailov, V.V., Bae, K.S. and Oh, H.W., 2006. Formosa agariphila sp. nov., a budding bacterium of the family Flavobacteriaceae isolated from marine environments, and emended description of the genus Formosa. International journal of systematic and evolutionary microbiology, 56(1), pp.161-167.

Needoba, J.A., Foster, R.A., Sakamoto, C., Zehr, J.P. and Johnson, K.S., 2007. Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. Limnology and Oceanography, 52(4), pp.1317-1327.

Neuwirth, E., 2011. RColorBrewer: ColorBrewer palettes. R package version, 1(5).

Niederberger, T.D., Sohm, J.A., Tirindelli, J., Gunderson, T., Capone, D.G., Carpenter, E.J. and Cary, S.C., 2012. Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys. FEMS microbiology ecology, 82(2), pp.376-390.

Noble, A.E., Lamborg, C.H., Ohnemus, D.C., Lam, P.J., Goepfert, T.J., Frame, C.H., Casciotti, K.L., DiTullio, G.R., Jennings, J.C. and Saito, M.A., 2012. Basin-scale inputs of cobalt, iron, and manganese from the Benguela-Angola front to the South Atlantic Ocean.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y. and Stepanauskas, R., 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. Journal of Computational Biology, 20(10), pp.714-737.

Orme-Johnson, W.H., 1985. Molecular basis of biological nitrogen fixation. Annual review of biophysics and biophysical chemistry, 14(1), pp.419-459.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R. and Fonstein, M., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research, 33(17), pp.5691-5702.

Paerl, H.W. and Prufert, L.E., 1987. Oxygen-poor microzones as potential sites of microbial N_2 fixation in nitrogen-depleted aerobic marine waters. Applied and Environmental Microbiology, 53(5), pp.1078-1087.

Pantoja, S., Repeta, D.J., Sachs, J.P. and Sigman, D.M., 2002. Stable isotope constraints on the nitrogen cycle of the Mediterranean Sea water column. Deep Sea Research Part I: Oceanographic Research Papers, 49(9), pp.1609-1621.

Parker, M.W., Pattus, F., Tucker, A.D. and Tsernoglou, D., 1989. Structure of the membrane-pore-forming fragment of colicin A.

Paulus, A., Rossius, S.G.H., Dijk, M. and de Vries, S., 2012. Oxoferryl-porphyrin radical catalytic intermediate in cytochrome bd oxidases protects cells from formation of reactive oxygen species. Journal of Biological Chemistry, 287(12), pp.8830-8838.

Piro, V.C., Faoro, H., Weiss, V.A., Steffens, M.B., Pedrosa, F.O., Souza, E.M. and Raittz, R.T., 2014. FGAP: an automated gap closing tool. BMC research notes, 7(1), p.371.

Ploug, H., Kühl, M., Buchholz-Cleven, B. and Jørgensen, B.B., 1997. Anoxic aggregates-an ephemeral phenomenon in the pelagic environment?. Aquatic Microbial Ecology, 13(3), pp.285-294.

Podschun, R. and Ullmann, U., 1998. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clinical microbiology reviews, 11(4), pp.589-603.

Pomeroy, L.R. and Wiebe, W.J., 2001. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. Aquatic Microbial Ecology, 23(2), pp.187-204.

Pommier, T., Pinhassi, J. and Hagström, Å., 2005. Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. Aquatic Microbial Ecology, 41(1), pp.79-89.

Poole, R.K. and Hill, S., 1997. Respiratory protection of nitrogenase activity in Azotobacter vinelandii—roles of the terminal oxidases. Bioscience reports, 17(3), pp.303-317.

Porsby, C.H., Nielsen, K.F. and Gram, L., 2008. Phaeobacter and Ruegeria species of the Roseobacter clade colonize separate niches in a Danish turbot (Scophthalmus maximus)-rearing farm and antagonize Vibrio anguillarum under different growth conditions. Applied and environmental microbiology, 74(23), pp.7356-7364.

Postgate, J.R., 1982. Biology nitrogen fixation: fundamentals. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 296(1082), pp.375-385.

Price, M.N., Dehal, P.S. and Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one, 5(3), p.e9490.

Price, P.B. and Sowers, T., 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. Proceedings of the National Academy of Sciences of the United States of America, 101(13), pp.4631-4636.

Pronk, J.T., Meulenberg, R., Hazeu, W., Bos, P. and Kuenen, J.G., 1990. Oxidation of reduced inorganic sulphur compounds by acidophilic thiobacilli. FEMS Microbiology letters, 75(2-3), pp.293-306.

Pryde, S.E., Duncan, S.H., Hold, G.L., Stewart, C.S. and Flint, H.J., 2002. The microbiology of butyrate formation in the human colon. FEMS microbiology letters, 217(2), pp.133-139.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing.

Raes, J., Letunic, I., Yamada, T., Jensen, L.J. and Bork, P., 2011. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. Molecular systems biology, 7(1), p.473.

Rahav, E., Bar-Zeev, E., Ohayion, S., Elifantz, H., Belkin, N., Herut, B., Mulholland, M.R. and Berman-Frank, I.R., 2013c. N_2 fixation in aphotic oxygenated marine environments. Frontiers in microbiology, 4, p.227.

Rahav, E., Herut, B., Mulholland, M.R., Belkin, N., Elifantz, H. and Berman-Frank, I., 2015. Heterotrophic and autotrophic contribution to N_2 fixation in the Gulf of Aqaba. Marine Ecology Progress Series, 522, pp.67-77.

Rahav, E., Herut, B., Mulholland, M.R., Voß, B., Stazic, D., Steglich, C., Hess, W.R. and Berman-Frank, I., 2013b. Contribution of N_2 fixation to bacterial and primary productivity in the Gulf of Aqaba (Red Sea). Biogeosciences Discussions, 10(6), pp.10327-10361.

Rahav, E., Herut, B., Stambler, N., Bar-Zeev, E., Mulholland, M.R. and Berman-Frank, I., 2013a. Uncoupling between N_2 fixation and primary productivity in the eastern Mediterranean Sea. Journal of Geophysical Research: Biogeosciences, 118(1), pp.195-202.

Raimbault, P. and Garcia, N., 2007. Carbon and nitrogen uptake in the South Pacific Ocean: evidence for efficient N_2 fixation and regenerated production leading to large accumulation of dissolved organic matter in nitrogen-depleted waters. Biogeosciences Discussions, 4(5), pp.3531-3579.

Ramakers, C., Ruijter, J. M., Deprez, R. H. L., Moorman, A. F., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters, 339(1), pp.62-66.

Ratten, J.M., LaRoche, J., Desai, D.K., Shelley, R.U., Landing, W.M., Boyle, E., Cutter, G.A. and Langlois, R.J., 2015. Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. Deep Sea Research Part II: Topical Studies in Oceanography, 116, pp.332-341.

Reddy, C.S.K., Ghai, R. and Kalia, V., 2003. Polyhydroxyalkanoates: an overview. Bioresource technology, 87(2), pp.137-146.

Reddy, K.J., Haskell, J.B., Sherman, D.M. and Sherman, L.A., 1993. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. Journal of Bacteriology, 175(5), pp.1284-1292.

Redfield, A.C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton (pp. 176-192). James Johnstone memorial volume: university press of liverpool.

Redfield, A.C., 1963. The influence of organisms on the composition of sea-water. The sea, pp.26-77.

Rees, A.P., Gilbert, J.A. and Kelly-Gerreyn, B.A., 2009. Nitrogen fixation in the western English Channel (NE Atlantic ocean). Marine Ecology Progress Series, 374, pp.7-12.

Riemann, L., Farnelid, H. and Steward, G.F., 2010. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. Aquatic Microbial Ecology, 61(3), pp.235-247.

Rijkenberg, M.J., Steigenberger, S., Powell, C.F., Haren, H., Patey, M.D., Baker, A.R. and Achterberg, E.P., 2012. Fluxes and distribution of dissolved iron in the eastern (sub-) tropical North Atlantic Ocean. Global Biogeochemical Cycles, 26(3).

Rubin, M., Berman-Frank, I., Shaked, Y., 2011. Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. Nature Geoscience, 4, pp.529–534.

Sabra, W., Zeng, A.P., Lünsdorf, H. and Deckwer, W.D., 2000. Effect of oxygen on formation and structure ofAzotobacter vinelandii alginate and its role in

protecting nitrogenase. Applied and environmental microbiology, 66(9), pp.4037-4044.

Sahu, S.N., Acharya, S., Tuminaro, H., Patel, I., Dudley, K., LeClerc, J.E., Cebula, T.A. and Mukhopadhyay, S., 2003. The bacterial adaptive response gene, barA, encodes a novel conserved histidine kinase regulatory switch for adaptation and modulation of metabolism in Escherichia coli. Molecular and cellular biochemistry, 253(1), pp.167-177.

Sargent, E.C., Hitchcock, A., Johansson, S.A., Langlois, R., Moore, C.M., LaRoche, J., Poulton, A.J. and Bibby, T.S., 2016. Evidence for polyploidy in the globally important diazotroph *Trichodesmium*. FEMS Microbiology Letters, 363(21), p.fnw244.

Scharek, R., Tupas, L.M. and Karl, D.M., 1999. Diatom fluxes to the deep sea in the oligotrophic North Pacific gyre at Station ALOHA. Marine Ecology Progress Series, 182, pp.55-67.

Schlesier, J., Rohde, M., Gerhardt, S. and Einsle, O., 2015. A Conformational Switch Triggers Nitrogenase Protection from Oxygen Damage by Shethna Protein II (FeSII). Journal of the American Chemical Society, 138(1), pp.239-247.

Schlitzer, R., 2012. Ocean data view.

Schlitzer, R., 2015. Ocean Data View. odv. awi. de.

Schneiker, S., dos Santos, V.A.M., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., Chernikova, T.N., Denaro, R., Ferrer, M., Gertler, C. and Goesmann, A., 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium Alcanivorax borkumensis. Nature biotechnology, 24(8), pp.997-1004.

Schunck, H., Lavik, G., Desai, D.K., Großkopf, T., Kalvelage, T., Löscher, C.R., Paulmier, A., Contreras, S., Siegel, H., Holtappels, M. and Rosenstiel, P., 2013. Giant hydrogen sulfide plume in the oxygen minimum zone off Peru supports chemolithoautotrophy. PLoS One, 8(8), p.e68661.

Severin, I., Bentzon-Tilia, M., Moisander, P.H. and Riemann, L., 2015. Nitrogenase expression in estuarine bacterioplankton influenced by organic carbon and availability of oxygen. FEMS microbiology letters, 362(14), p.fnv105.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), pp.2498-2504.

Shelley, R.U., Morton, P.L. and Landing, W.M., 2015. Elemental ratios and enrichment factors in aerosols from the US-GEOTRACES North Atlantic transects. Deep Sea Research Part II: Topical Studies in Oceanography, 116, pp.262-272.

Shieh, W.Y., Simidu, U. and Maruyama, Y., 1989. Enumeration and characterization of nitrogen-fixing bacteria in an eelgrass (Zostera marina) bed. Microbial ecology, 18(3), pp.249-259.

Shiller, A. M., 1997. Manganese in surface waters of the Atlantic Ocean. Geophysical Research Letters, 24(1), pp.1495-1498.

Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S. and Furuya, K., 2014. Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean. Global Biogeochemical Cycles, 28(10), pp.1096-1110.

Sieburth, J.M., 1987. Contrary habitats for redox-specific processes: methanogenesis in oxic waters and oxidation in anoxic waters. Microbes in the Sea, pp.11-38.

Simidu, U., Kita-Tsukamoto, K.U.M.I.K.O., YASUMOTO, T. and YOTSU, M., 1990. Taxonomy of four marine bacterial strains that produce tetrodotoxin. International Journal of Systematic and Evolutionary Microbiology, 40(4), pp.331-336.

Singh, B.K. and Walker, A., 2006. Microbial degradation of organophosphorus compounds. FEMS microbiology reviews, 30(3), pp.428-471.

Smith, D.C., Simon, M., Alldredge, A.L. and Azam, F., 1992. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. Nature, 359(6391), pp.139-142.

Sobarzo, M., Bravo, L., Donoso, D., Garcés-Vargas, J. and Schneider, W., 2007. Coastal upwelling and seasonal cycles that influence the water column over the continental shelf off central Chile. Progress in Oceanography, 75(3), pp.363-382.

Sohm, J.A., Hilton, J.A., Noble, A.E., Zehr, J.P., Saito, M.A. and Webb, E.A., 2011a. Nitrogen fixation in the South Atlantic Gyre and the Benguela upwelling system. Geophysical Research Letters, 38(16).

Sohm, J.A., Subramaniam, A., Gunderson, T.E., Carpenter, E.J. and Capone, D.G., 2011b. Nitrogen fixation by *Trichodesmium* spp. and unicellular diazotrophs in the North Pacific Subtropical Gyre. Journal of Geophysical Research: Biogeosciences, 116(G3).

Spietz, R.L., Williams, C.M., Rocap, G. and Horner-Devine, M.C., 2015. A Dissolved Oxygen Threshold for Shifts in Bacterial Community Structure in a Seasonally Hypoxic Estuary. PloS one, 10(8), p.e0135731.

Stal, L. J., 2009. Is the distribution of nitrogen-fixing cyanobacteria in the oceans related to temperature? Environmental Microbiology, 11(7), pp.1632-1645.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), pp.1312-1313.

Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.E.T., Sachdeva, R., Jones, A.C., Schwalbach, M.S. and Rose, J.M., 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME journal, 5(9), pp.1414-1425.

Stingl, U., Desiderio, R.A., Cho, J.C., Vergin, K.L. and Giovannoni, S.J., 2007. The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. Applied and environmental microbiology, 73(7), pp.2290-2296.

Stoffer, D., 2012. astsa: Applied Statistical Time Series Analysis. R package version, 1.

Stramma, L., Johnson, G. C., Sprintall, J., Mohrholz, V., 2008. Expanding oxygen-minimum zones in the tropical oceans. Science, 320(5876), pp.655-658.

Stramma, L., Prince, E.D., Schmidtko, S., Luo, J., Hoolihan, J.P., Visbeck, M., Wallace, D.W., Brandt, P. and Körtzinger, A., 2012. Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. Nature Climate Change, 2(1), pp.33-37.

Strom, S.L., 2008. Microbial ecology of ocean biogeochemistry: a community perspective. Science, 320(5879), pp.1043-1045.

Subramaniam, A., Mahaffey, C., Johns, W. and Mahowald, N., 2013. Equatorial upwelling enhances nitrogen fixation in the Atlantic Ocean. Geophysical Research Letters, 40(9), pp.1766-1771.

Subramaniam, A., Yager, P. L., Carpenter, E. J., Mahaffey, C., Bjorkman, K., Cooley, S., Capone, D. G., 2008. Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. Proceedings of the National Academy of Sciences of the United States of America, 105(30), pp.10460-10465.

Summons, R.E., Jahnke, L.L., Hope, J.M. and Logan, G.A., 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature, 400(6744), pp.554-557.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. and Cornejo-Castillo, F.M., 2015. Structure and function of the global ocean microbiome. Science, 348(6237), p.1261359.

Sunamura, M., Higashi, Y., Miyako, C., Ishibashi, J.I. and Maruyama, A., 2004. Two bacteria phylotypes are predominant in the Suiyo Seamount hydrothermal plume. Applied and Environmental Microbiology, 70(2), pp.1190-1198.

Suttle, C.A., 2007. Marine viruses—major players in the global ecosystem. Nature Reviews Microbiology, 5(10), pp.801-812.

Suyama, M., Torrents, D. and Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research, 34(suppl 2), pp.W609-W612.

Suzuki, M.T. and Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and environmental microbiology, 62(2), pp.625-630.

Swannell, R.P., Lee, K. and McDonagh, M., 1996. Field evaluations of marine oil spill bioremediation. Microbiological Reviews, 60(2), pp.342-365.

Tamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution, 24(8), pp.1596-1599.

Tay, K.L., Doe, K.G., Wade, S.J., Vaughan, D.A., Berrigan, R.E. and Moore, M.J., 1992. Sediment bioassessment in Halifax harbour. Environmental toxicology and chemistry, 11(11), pp.1567-1581.

Team, R.C., 2013. R: A language and environment for statistical computing.

Team, R.C., 2015. R: A language and environment for statistical computing.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., Kassabgy, M., Huang, S., Mann, A.J., Waldmann, J. and Weber, M.,

2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science, 336(6081), pp.608-611.

Thamdrup, B., Dalsgaard, T. and Revsbech, N.P., 2012. Widespread functional anoxia in the oxygen minimum zone of the Eastern South Pacific. Deep Sea Research Part I: Oceanographic Research Papers, 65, pp.36-45.

Thompson, A. W., Zehr, J. P., 2013. Cellular interactions: Lessons from the nitrogen-fixing cyanobacteria. Journal of Phycology, 49(6), pp.1024-1035.

Thompson, A., Carter, B.J., Turk-Kubo, K., Malfatti, F., Azam, F. and Zehr, J.P., 2014. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. Environmental microbiology, 16(10), pp.3238-3249.

Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaulot, D., Kuypers, M.M. and Zehr, J.P., 2012. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. Science, 337(6101), pp.1546-1550.

Thompson, R.C., Moore, C.J., Vom Saal, F.S. and Swan, S.H., 2009. Plastics, the environment and human health: current consensus and future trends. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1526), pp.2153-2166.

Thorneley, R.N. and Ashby, G.A., 1989. Oxidation of nitrogenase iron protein by dioxygen without inactivation could contribute to high respiration rates of Azotobacter species and facilitate nitrogen fixation in other aerobic environments. Biochemical Journal, 261(1), pp.181-187.

Toulza, E., Tagliabue, A., Blain, S. and Piganeau, G., 2012. Analysis of the global ocean sampling (GOS) project for trends in iron uptake by surface ocean microbes. PLoS One, 7(2), p.e30931.

Tritt, A., Eisen, J.A., Facciotti, M.T. and Darling, A.E., 2012. An integrated pipeline for de novo assembly of microbial genomes. PloS one, 7(9), p.e42304.

Tu, Q., Deng, Y., Yan, Q., Shen, L., Lin, L., He, Z., Wu, L., Van Nostrand, J.D., Buzzard, V., Michaletz, S.T. and Enquist, B.J., 2016. Biogeographic patterns of soil diazotrophic communities across six forests in the North America. Molecular ecology.

Turk, K. A., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., Gilbert, J., 2011. Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. The ISME Journal, 5(7), pp.1201-1212.

Turk-Kubo, K.A., Frank, I.E., Hogan, M.E., Desnues, A., Bonnet, S. and Zehr, J.P., 2015. Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon). Biogeosciences, 12(24), pp.7435-7452.

Turk-Kubo, K.A., Karamchandani, M., Capone, D.G. and Zehr, J.P., 2014. The paradox of marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. Environmental microbiology, 16(10), pp.3095-3114.

Turnbaugh, P.J., Biomolecules, S.B.D. and Roscoff, F., 2011. Environmental and gut bacteroidetes: the food connection. Human health and disease in a microbial world, p.96.

Uchida, M., Nakata, K. and Maeda, M., 1997. Conversion of Ulva fronds to a hatchery diet for Artemia nauplii utilizing the degrading and attaching abilities of Pseudoalteromonas espejiana. Journal of applied phycology, 9(6), pp.541-549.

Uchida, M., Nakayama, A. and Abe, S., 1995. Distribution and characterization of bacteria capable of decomposing brown algae fronds in waters associated with Laminaria vegetation. Fisheries science, 61(1), pp.117-120.

Udiković-Kolić, N., Scott, C. and Martin-Laurent, F., 2012. Evolution of atrazine-degrading capabilities in the environment. Applied microbiology and biotechnology, 96(5), pp.1175-1189.

Van Beilen, J.B., Marin, M.M., Smits, T.H., Röthlisberger, M., Franchini, A.G., Witholt, B. and Rojo, F., 2004. Characterization of two alkane hydroxylase genes from the marine hydrocarbonoclastic bacterium Alcanivorax borkumensis. Environmental microbiology, 6(3), pp.264-273.

van der Maarel, M.J., Sprenger, W., Haanstra, R. and Forney, L.J., 1999. Detection of methanogenic archaea in seawater particles and the digestive tract of a marine fish species. FEMS microbiology letters, 173(1), pp.189-194.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P.C., Colwell, R.R., Höfle, M.G. and Pruzzo, C., 2012. Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. The ISME journal, 6(1), pp.21-30.

Viana, M., Querol, X., Alastuey, A., Cuevas, E., Rodrıguez, S., 2002. Influence of African dust on the levels of atmospheric particulates in the Canary Islands air quality network. Atmospheric Environment, 36(38), pp.5861-5875.

Villareal, T.A., 1990. Laboratory culture and preliminary characterization of the nitrogen-fixing Rhizosolenia-Richelia symbiosis. Marine Ecology, 11(2), pp.117-132.

Vitousek, P. M., Howarth, R. W., 1991. Nitrogen limitation on land and in the sea: How can it occur? Biogeochemistry, 13(2), pp.87-115.

Vitousek, P.M., Menge, D.N., Reed, S.C. and Cleveland, C.C., 2013. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1621), p.20130119.

Vojvodic, A., Medford, A.J., Studt, F., Abild-Pedersen, F., Khan, T.S., Bligaard, T. and Nørskov, J.K., 2014. Exploring the limits: A low-pressure, low-temperature Haber–Bosch process. Chemical Physics Letters, 598, pp.108-112.

Vollmers, J., Voget, S., Dietrich, S., Gollnow, K., Smits, M., Meyer, K., Brinkhoff, T., Simon, M. and Daniel, R., 2013. Poles apart: Arctic and Antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. PLoS One, 8(5), p.e63422.

Voss, M., Bombar, D., Loick, N. and Dippner, J.W., 2006. Riverine influence on nitrogen fixation in the upwelling region off Vietnam, South China Sea. Geophysical Research Letters, 33(7).

Voss, M., Croot, P., Lochte, K., Mills, M. and Peeken, I., 2004. Patterns of nitrogen fixation along 10 N in the tropical Atlantic. Geophysical Research Letters, 31(23).

Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, S.G., Tortell, P.D. and Hallam, S.J., 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science, 326(5952), pp.578-582.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology, 73(16), pp.5261-5267.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S. and Schwartz, M., 2013. gplots: Various R programming tools for plotting data. R package version 2.12. 1.

Wasmund, N., Struck, U., Hansen, A., Flohr, A., Nausch, G., Grüttmüller, A. and Voss, M., 2015. Missing nitrogen fixation in the Benguela region. Deep Sea Research Part I: Oceanographic Research Papers, 106, pp.30-41.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R. and Ley, R.E., 2012. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME journal, 6(1), pp.94-103.

Wickham, H. and Chang, W., 2015. Ggplot2: An implementation of the grammar of graphics, Version 1.0. 1.

Williams, K.P., Gillespie, J.J., Sobral, B.W., Nordberg, E.K., Snyder, E.E., Shallom, J.M. and Dickerman, A.W., 2010. Phylogeny of gammaproteobacteria. Journal of bacteriology, 192(9), pp.2305-2314.

Wintzingerode, F.V., Göbel, U.B. and Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS microbiology reviews, 21(3), pp.213-229.

Witty, J.F. and Minchin, F.R., 1998. Hydrogen measurements provide direct evidence for a variable physical barrier to gas diffusion in legume nodules. Journal of Experimental Botany, 49(323), pp.1015-1020.

Wozniak, A. S., Shelley, R. U., Sleighter, R. L., Abdulla, H. A. N., Morton, P. L., Landing, W. M. and Hatcher, P. G., 2013. Relationships among aerosol water soluble organic matter, iron and aluminum in European, North African, and Marine air masses from the 2010 US GEOTRACES cruise. Marine Chemistry, 154, pp.24-33.

Wright, J.J., Konwar, K.M. and Hallam, S.J., 2012. Microbial ecology of expanding oxygen minimum zones. Nature Reviews Microbiology, 10(6), pp.381-394.

Wurl, O., Zimmer, L., Cutter, G. A., 2013. Arsenic and phosphorus biogeochemistry in the ocean: Arsenic species as proxies for P-limitation. Limnology and. Oceanography, 58(2), pp.729-740.

Xiao, P., Jiang, Y., Liu, Y., Tan, W., Li, W. and Li, R., 2015. Re-evaluation of the diversity and distribution of diazotrophs in the South China Sea by pyrosequencing the nifH gene. Marine and Freshwater Research, 66(8), pp.681-691.

Yakimov, M.M., Giuliano, L., Denaro, R., Crisafi, E., Chernikova, T.N., Abraham, W.R., Luensdorf, H., Timmis, K.N. and Golyshin, P.N., 2004. Thalassolituus oleivorans gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. International Journal of Systematic and Evolutionary Microbiology, 54(1), pp.141-148.

Yakimov, M.M., Timmis, K.N. and Golyshin, P.N., 2007. Obligate oil-degrading marine bacteria. Current opinion in biotechnology, 18(3), pp.257-266.

Yamada, K.D., Tomii, K. and Katoh, K., 2016. Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. Bioinformatics, p.btw412.

Yogev, T., Rahav, E., Bar-Zeev, E., Man-Aharonovich, D., Stambler, N., Kress, N., Béjà, O., Mulholland, M.R., Herut, B. and Berman-Frank, I., 2011. Is N_2 fixation significant in the Levantine Basin, east Mediterranean Sea?. Environmental microbiology, 13(4), pp.854-871.

Yoon, J.H., Kang, S.J. and Oh, T.K., 2006. Polaribacter dokdonensis sp. nov., isolated from seawater. International journal of systematic and evolutionary microbiology, 56(6), pp.1251-1255.

Zehr, J.P., Mellon, M.T. and Zani, S., 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes. Applied and environmental microbiology, 64(9), pp.3444-3450.

Zehr, J.P. and Kudela, R.M., 2011. Nitrogen cycle of the open ocean: from genes to ecosystems. Annual Review of Marine Science, 3, pp.197-225.

Zehr, J.P., 2011. Nitrogen fixation by marine cyanobacteria. Trends in microbiology, 19(4), pp.162-173.

Zehr, J.P. and Turner, P.J., 2001. Nitrogen fixation: nitrogenase genes and gene expression. Methods in microbiology, 30, pp.271-286.

Zehr, J.P., Bench, S.R., Carter, B.J., Hewson, I., Niazi, F., Shi, T., Tripp, H.J. and Affourtit, J.P., 2008. Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. Science, 322(5904), pp.1110-1112.

Zehr, J.P., Crumbliss, L.L., Church, M.J., Omoregie, E.O. and Jenkins, B.D., 2003b. Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of functional genes. Biotechniques, 35(5), pp.996-1013.

Zehr, J.P., Jenkins, B.D., Short, S.M. and Steward, G.F., 2003a. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental microbiology, 5(7), pp.539-554.

Zehr, J.P., Montoya, J.P., Jenkins, B.D., Hewson, I., Mondragon, E., Short, C.M., Church, M.J., Hansen, A. and Karl, D.M., 2007. Experiments linking nitrogenase gene expression to nitrogen fixation in the North Pacific subtropical gyre. Limnology and Oceanography, 52(1), pp.169-183.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), pp.614-620.

Zhao, G. and Winkler, M.E., 1996. A novel alpha-ketoglutarate reductase activity of the serA-encoded 3-phosphoglycerate dehydrogenase of Escherichia coli K-

12 and its possible implications for human 2-hydroxyglutaric aciduria. Journal of bacteriology, 178(1), pp.232-239.

Zimmer, L.A. and G. A. Cutter., 2012. High resolution determination of nanomolar concentrations of dissolved reactive phosphate in ocean surface waters using long path liquid waveguide capillary cells (LWCC) and spectrometric detection. Limnology and Oceanography: Methods, 10, pp.568-580.

APPENDIX A: SUPPLEMENTAL TABLES

Supplemental Table 1: BCO-DMO sets included in the PRIMER-E analysis.

Title	Website
GT10/11 - CTD - GT-C Bottle	http://www.bco-dmo.org/dataset/3517
	http://www.bco-dmo.org/dataset/3687
GT10/11 - CTD - GT-C Cast Sheets	http://www.bco-dmo.org/dataset/3508
	http://www.bco-dmo.org/dataset/3671
GT10/11 - CTD - GT-C Profiles[2)]	http://www.bco-dmo.org/dataset/3516
	http://www.bco-dmo.org/dataset/3698
GT10/11 - CTD - ODF/SIOR Bottle	http://www.bco-dmo.org/dataset/3519
	http://www.bco-dmo.org/dataset/3662
GT11 - CTD - ODF/SIOR Profiles	http://www.bco-dmo.org/dataset/3518
	http://www.bco-dmo.org/dataset/3699
GT10/11 - FIA-AlFe	http://www.bco-dmo.org/dataset/3821
	http://www.bco-dmo.org/dataset/3822
GT10/11 - Nanomolar Nutrients – Profiles	http://www.bco-dmo.org/dataset/3521
	http://www.bco-dmo.org/dataset/3839
GT10/11 - Nanomolar Nutrients – Surface	http://www.bco-dmo.org/dataset/3470
	http://www.bco-dmo.org/dataset/3838
GT10-11 - Co Total Dissolved and Labile	http://www.bco-dmo.org/dataset/3868
GT10-11 - dFe and dFe(II)	http://www.bco-dmo.org/dataset/3826
GT10-11 - Fe Mn Zn Cd and Cu	http://www.bco-dmo.org/dataset/3861
GT10-11 - Filtered Rainwater	http://www.bco-dmo.org/dataset/3863
GT10-11 - GaPbBa_bottles	http://www.bco-dmo.org/dataset/3827
GT10-11 - GaPbBa_surface	http://www.bco-dmo.org/dataset/3831
GT10-11 - Hg Aerosol	http://www.bco-dmo.org/dataset/3854
GT10-11 - Hg Particulate	http://www.bco-dmo.org/dataset/3859
GT10-11 - Hg Speciation	http://www.bco-dmo.org/dataset/3860
GT10-11 - N African Mineral Dust Composition	http://www.bco-dmo.org/dataset/4063
GT10-11 - Trace Metals Aerosol	http://www.bco-dmo.org/dataset/3865
GT10-11 - Unfiltered Rainwater	http://www.bco-dmo.org/dataset/3858

1) Data sets shaded grey were included in the PRIMER-E analysis, but did not contribute significantly to results.
2) Data sets without shading either led to significant results during the PRIMER-E analysis, or are included in the reported results because of previously established significance in the literature.
3) The particulate composition was analysed in a separate PRIMER-E matrix to minimize the number of samples deleted due to missing values.

Supplemental Table 2: Results of the BEST analysis for the three matrices: Entire dataset, SML subset and deep samples subset.

	Entire dataset[1)]	**SML samples**[2)]	**Deep samples**[3)]
Significance[4)]	1%	1%	31%
Rho[4)]	0.370	0.622	0.192
Variables[5)]	Longitude	Longitude	Latitude
	MLD	Latitude	Depth
	SiO_2	NO_2^-	Dissolved Fe
	N:P	N:P	Mn
	Dissolved Fe	Dissolved Al	Pb
	Temperature	Dissolved Fe	
	Pb		

1) This matrix included all data points after initial preparation as described in the methods.
2) This matrix contains the subset of samples from the entire dataset located in the SML.
3) This matrix contains the subset of samples from the entire dataset located below the SML.
4) A Rho value with a significance level lower than 5% indicates that variables correlate significantly with *nifH* phylotype abundances.
5) Listed are the variables that correlate best with variations in *nifH* phylotype abundances.

Supplemental Table 3: Samples collected for *nifH* high-throughput Tag-sequencing.

Sample Name	Cruise	Vessel	Longitude (°North)	Latitude (°East)	Depth (m)	Date (dd/mm/yyyy)	Temperature CTD (°C)
71	ANTXXVI-I	Polarstern	-32.9260	-26.1474	1	13/11/2009	24.34
72	ANTXXVI-I	Polarstern	-34.1656	-26.6672	1	14/11/2009	24.32
73	ANTXXVI-I	Polarstern	-34.7076	-27.1235	1	14/11/2009	24.12
74	ANTXXVI-I	Polarstern	-35.8216	-28.1224	1	14/11/2009	24.11
75	ANTXXVI-I	Polarstern	-36.7725	-28.9680	1	15/11/2009	23.03
76	ANTXXVI-I	Polarstern	-37.8380	-29.9076	1	15/11/2009	21.53
77	ANTXXVI-I	Polarstern	-38.7152	-30.6744	1	15/11/2009	21.17
78	ANTXXVI-I	Polarstern	-39.3630	-31.2343	1	16/11/2009	21.22
79	ANTXXVI-I	Polarstern	-40.4650	-32.1232	1	16/11/2009	20.94
80	ANTXXVI-I	Polarstern	-41.3998	-32.8708	1	16/11/2009	19.67
81	ANTXXVI-I	Polarstern	-42.4924	-33.7363	1	17/11/2009	19.83
82	ANTXXVI-I	Polarstern	-43.5601	-34.5741	1	17/11/2009	17.24
83	ANTXXVI-I	Polarstern	-44.4951	-35.3005	1	17/11/2009	17.67
84	ANTXXVI-I	Polarstern	-45.7754	-36.2851	1	18/11/2009	18.41
85	ANTXXVI-I	Polarstern	-46.9660	-37.1897	1	18/11/2009	18.70
86	ANTXXVI-I	Polarstern	-48.1864	-38.1057	1	18/11/2009	16.09
87	ANTXXVI-I	Polarstern	-49.4093	-39.0123	1	19/11/2009	17.31
88	ANTXXVI-I	Polarstern	-50.5239	-39.8379	1	19/11/2009	17.61
89	ANTXXVI-I	Polarstern	-50.9597	-40.1667	1	19/11/2009	17.48
AZMP.F.BBL07.020m	HUD2014030	Hudson	-65.349448	41.866181	20	0	21.0304
AZMP.F.CSL4.001m	HUD2014030	Hudson	-59.783753	47.270352	1	0	9.3775
AZMP.F.GULD04.001m	HUD2014030	Hudson	-58.900818	43.790474	1	0	16.5866
AZMP.F.GULD04.100m	HUD2014030	Hudson	-58.900818	43.790474	100	0	6.3971
AZMP.F.GULD04.250m	HUD2014030	Hudson	-58.900818	43.790474	250	0	10.8165
AZMP.F.HL02.001m	HUD2014030	Hudson	-63.316795	44.266707	1	0	17.6059
AZMP.F.HL02.080m	HUD2014030	Hudson	-63.316795	44.266707	80	0	4.4029
AZMP.F.HL08.001m	HUD2014030	Hudson	-61.344907	42.36328	1	0	20.9704
AZMP.F.HL08.250m	HUD2014030	Hudson	-61.344907	42.36328	250	0	12.0326
AZMP.F.STAB01.001m	HUD2014030	Hudson	-59.528857	45.998645	1	0	14.1255
AZMP.F.STAB01.020m	HUD2014030	Hudson	-59.528857	45.998645	20	0	13.6116
AZMP.S.BBL1.001m	HUD2014004	Hudson	-65.480256	43.249548	1	0	1.7595
AZMP.S.BBL1.040m	HUD2014004	Hudson	-65.480256	43.249548	40	0	1.9476
AZMP.S.BBL7.250m	HUD2014004	Hudson	-65.349448	41.866181	250	0	10.7695
AZMP.S.CLS04.001m	HUD2014004	Hudson	-59.783753	47.270352	1	0	-0.4311
AZMP.S.GULD04.001m	HUD2014004	Hudson	-58.900818	43.790474	1	0	1.7715
AZMP.S.GULD04.100m	HUD2014004	Hudson	-58.900818	43.790474	100	0	2.6922
AZMP.S.GULD04.250m	HUD2014004	Hudson	-58.900818	43.790474	250	0	11.4902
AZMP.S.HL02.001m	HUD2014004	Hudson	-63.316795	44.266707	1	0	-0.5783
AZMP.S.HL02.080m	HUD2014004	Hudson	-63.316795	44.266707	80	0	-0.4019
AZMP.S.HL08.250m	HUD2014004	Hudson	-61.344907	42.36328	250	0	10.1657
AZMP.S.STAB01.001m	HUD2014004	Hudson	-59.528857	45.998645	1	0	-0.1209
AZMP.S.STAB01.020m	HUD2014004	Hudson	-59.528857	45.998645	20	0	-0.3912

BB.17Dec2014.01m	Bedford Basin	SigmaT	-63.64	44.69	1	17/12/2014	5.9
BB.17Dec2014.10m	Bedford Basin	SigmaT	-63.64	44.69	10	17/12/2014	6.09
BB.17Dec2014.60m	Bedford Basin	SigmaT	-63.64	44.69	60	17/12/2014	4.95
BB.18Jun2014.01m	Bedford Basin	SigmaT	-63.64	44.69	1	18/06/2014	11.44
BB.18Jun2014.10m	Bedford Basin	SigmaT	-63.64	44.69	10	18/06/2014	8.4
BB.18Jun2014.60m	Bedford Basin	SigmaT	-63.64	44.69	60	18/06/2014	1.58
BB.19Mar2014.05m	Bedford Basin	SigmaT	-63.64	44.69	5	19/03/2014	1.23
BB.19Mar2014.60m	Bedford Basin	SigmaT	-63.64	44.69	60	19/03/2014	1.35
BB.24Sep2014.10m	Bedford Basin	SigmaT	-63.64	44.69	10	24/09/2014	14.53
BB.24Sep2014.60m	Bedford Basin	SigmaT	-63.64	44.69	60	24/09/2014	3.92
D361_C1_0	D361	RRS Discovery	-16.121759	28.312577	2.09735		19.754398
D361_C1_10	D361	RRS Discovery	-16.121759	28.312577	10.951134		19.746035
D361_C1_20	D361	RRS Discovery	-16.121759	28.312577	21.232813		19.744391
D361_C1_60	D361	RRS Discovery	-16.121759	28.312577			
D361_C1_80	D361	RRS Discovery	-16.121759	28.312577	78.826218		19.633625
D361_C15_30	D361	RRS Discovery	-21.817872	12.577107			
D361_C5_30	D361	RRS Discovery	-17.913264	12.587028	30.635984		19.18856
D361_C9_20	D361	RRS Discovery	-17.56857	12.586956	20.603694		18.909658
D361_C9_40	D361	RRS Discovery	-17.56857	12.586956	40.580557		17.449692
D361_C9_85	D361	RRS Discovery	-17.56857	12.586956	84.742285		16.307589
D361_F01	D361	RRS Discovery	-18.88444444	26.98222222	5	2011-02-08T08:06:00	20.569957
D361_F02	D361	RRS Discovery	-18.64277778	27.04055556	5	2011-02-08T18:23:00	20.247231
D361_F04	D361	RRS Discovery	-16.452421	27.03844	5	2011-02-18T16:02:00	20.834813
D361_F05	D361	RRS Discovery	-16.534938	26.682622	5	2011-02-18T18:00:00	20.782655
D361_F08	D361	RRS Discovery	-16.843049	25.614197	5	2011-02-19T00:00:00	20.418032
D361_F09	D361	RRS Discovery	-16.941918	25.265379	5	2011-02-19T02:00:00	19.998852
D361_F10	D361	RRS Discovery	-17.032483	24.922665	5	2011-02-19T03:55:00	20.414186
D361_F11	D361	RRS Discovery	-17.181669	24.45493	5	2011-02-19T06:30:00	20.174719
D361_F12	D361	RRS Discovery	-17.249033	24.179815	5	2011-02-19T08:00:00	20.255972
D361_F13	D361	RRS Discovery	-17.336804	23.848127	5	2011-02-19T10:05:00	20.223291
D361_F15	D361	RRS Discovery	-17.526322	23.157319	5	2011-02-19T14:02:00	19.780557
D361_F18	D361	RRS Discovery	-17.819048	22.071608	5	2011-02-19T20:05:00	18.585478
D361_F23	D361	RRS Discovery	-18.15758	19.891229	5	2011-02-20T08:03:00	20.993249
D361_F24	D361	RRS Discovery	-18.148953	19.604498	5	2011-02-20T10:05:00	20.185683
D361_F25	D361	RRS Discovery	-18.156667	19.287388	5	2011-02-20T11:54:00	20.584301
D361_F26	D361	RRS Discovery	-18.160812	18.896768	5	2011-02-20T14:05:00	19.878584
D361_F27	D361	RRS Discovery	-18.15561	18.546448	5	2011-02-20T16:00:00	20.175531
D361_F29	D361	RRS Discovery	-18.154391	17.864501	5	2011-02-20T20:00:00	20.257116
D361_F30	D361	RRS Discovery	-18.16294	17.478672	5	2011-02-20T22:15:00	22.114305
D361_F31	D361	RRS Discovery	-18.153068	16.989359	5	2011-02-21T01:00:00	21.663504
D361_F32	D361	RRS Discovery	-18.148106	16.607263	5	2011-02-21T03:00:00	22.248659
D361_F33	D361	RRS Discovery	-18.174851	16.226136	5	2011-02-21T04:55:00	19.465845
D361_F34	D361	RRS Discovery	-18.160073	15.79818	5	2011-02-21T07:00:00	21.178603
D361_F35	D361	RRS Discovery	-18.165556	15.380595	5	2011-02-21T09:02:00	21.420897
D361_F36	D361	RRS Discovery	-18.140021	15.075235	5	2011-02-21T11:05:00	19.502722
D361_F37	D361	RRS Discovery	-18.155355	14.72951	5	2011-02-21T12:53:00	19.507143
D361_F49	D361	RRS Discovery	-17.948874	12.585315	5	2011-02-23T23:50:00	22.455381

D361_F55	D361	RRS Discovery	-19.165412	12.593601	5	2011-02-23T15:02:00	24.151312
D361_F56	D361	RRS Discovery	-19.522648	12.58613	5	2011-02-23T17:03:00	24.075353
D361_F57	D361	RRS Discovery	-19.874666	12.586138	5	2011-02-23T19:00:00	24.245113
Kn199_10_101	Kn199	RV Knorr	-20.8	17.4	101	10/21/2010	14.5076
Kn199_10_2	Kn199	RV Knorr	-20.8	17.4	2	10/21/2010	27.4
Kn199_10_234.5	Kn199	RV Knorr	-20.8	17.4	235	10/21/2010	12.6871
Kn199_10_27.2	Kn199	RV Knorr	-20.8	17.4	27	10/21/2010	26.1589
Kn199_10_416.6	Kn199	RV Knorr	-20.8	17.4	417	10/24/2010	10.8325
Kn199_10_51.8	Kn199	RV Knorr	-20.8	17.4	52	10/23/2010	17.9374
Kn199_5_111.6	Kn199	RV Knorr	-22.0	31.0	112	10/23/2010	18.5523
Kn199_5_2	Kn199	RV Knorr	-22.0	31.0	2	10/23/2010	24.7
Kn199_5_31.8	Kn199	RV Knorr	-22.0	31.0	32	10/27/2010	24.3331
Kn199_5_62.2	Kn199	RV Knorr	-22.0	31.0	62	10/28/2010	22.303
Kn199_7_151.8	Kn199	RV Knorr	-22.0	24.0	152	10/28/2010	17.2727
Kn199_7_2	Kn199	RV Knorr	-22.0	24.0	2	10/28/2010	25.7
Kn199_7_33.2	Kn199	RV Knorr	-22.0	24.0	33	10/28/2010	25.6623
Kn199_7_72.9	Kn199	RV Knorr	-22.0	24.0	73	10/30/2010	21.55
Kn199_9_106.4	Kn199	RV Knorr	-18.3	17.4	106	10/30/2010	14.7667
Kn199_9_2	Kn199	RV Knorr	-18.3	17.4	2	10/30/2010	28
Kn199_9_22.4	Kn199	RV Knorr	-18.3	17.4	22	10/30/2010	27.8796
Kn199_9_249.3	Kn199	RV Knorr	-18.3	17.4	249	10/30/2010	12.6629
Kn199_9_52.1	Kn199	RV Knorr	-18.3	17.4	52	10/30/2010	18.9721
Kn204_10_2	Kn204	RV Knorr	-64.1786	31.7565	2	22/11/2011	
Kn204_10_41	Kn204	RV Knorr	-64.1786	31.7565	41	19/11/2011	24.134
Kn204_10_75	Kn204	RV Knorr	-64.1786	31.7565	75	19/11/2011	24.1885
Kn204_10_800	Kn204	RV Knorr	-64.1786	31.7565	800	19/11/2011	10.5304
Kn204_13_100	Kn204	RV Knorr	-53.2258	28.6441	100	25/11/2011	20.6064
Kn204_13_235	Kn204	RV Knorr	-53.2258	28.6441	235	25/11/2011	17.9294
Kn204_13_45	Kn204	RV Knorr	-53.2258	28.6441	45	25/11/2011	25.1845
Kn204_13_850	Kn204	RV Knorr	-53.2258	28.6441	850	25/11/2011	
Kn204_16_111	Kn204	RV Knorr	-44.8262	26.1369	111	30/11/2011	21.9954
Kn204_16_2	Kn204	RV Knorr	-44.8262	26.1369	2	30/11/2011	25.3
Kn204_16_285	Kn204	RV Knorr	-44.8262	26.1369	285	30/11/2011	
Kn204_16_40	Kn204	RV Knorr	-44.8262	26.1369	40	30/11/2011	25.1372
Kn204_16_802	Kn204	RV Knorr	-44.8262	26.1369	802	30/11/2011	8.9283
Kn204_21_100	Kn204	RV Knorr	-32.6247	20.8839	100	06/11/2011	22.2037
Kn204_21_2	Kn204	RV Knorr	-32.6247	20.8839	2	06/11/2011	25.3682
Kn204_21_286	Kn204	RV Knorr	-32.6247	20.8839	286	06/11/2011	15.3895
Kn204_21_40	Kn204	RV Knorr	-32.6247	20.8839	40	06/11/2011	25.2204
Kn204_21_60	Kn204	RV Knorr	-32.6247	20.8839	60	06/11/2011	25.0135
Kn204_21_625	Kn204	RV Knorr	-32.6247	20.8839	625	06/11/2011	9.6561
Kn204_24_2	Kn204	RV Knorr	-24.5	17.4	2	09/11/2011	24.7
Kn204_24_286	Kn204	RV Knorr	-24.5	17.4	286	10/11/2011	
Kn204_24_38	Kn204	RV Knorr	-24.5	17.4	38	10/11/2011	24.6764
Kn204_24_430	Kn204	RV Knorr	-24.5	17.4	430	10/11/2011	10.7319
Kn204_24_48	Kn204	RV Knorr	-24.5	17.4	48	10/11/2011	24.6776
Kn204_24_72	Kn204	RV Knorr	-24.5	17.4	72	10/11/2011	20.1988

nifH.G.01.170m.A	GEOVIDE	Pourquoi Pas	-10.036	40.333	170	5/19/2023	12.6452
nifH.G.01.200m.C	GEOVIDE	Pourquoi Pas	-10.036	40.333	200	5/19/2025	12.5848
nifH.G.02.100m.C	GEOVIDE	Pourquoi Pas	-9.459	40.333	100	5/20/2018	13.3251
nifH.G.04.035m.A	GEOVIDE	Pourquoi Pas	-9.767	40.333	35	5/21/2016	14.1129
nifH.G.04.100m.B	GEOVIDE	Pourquoi Pas	-9.767	40.333	100	5/21/2018	13.0532
nifH.G.11.054m.C	GEOVIDE	Pourquoi Pas	-12.219	40.333	54	5/23/2018	13.6975
nifH.G.11.200m.A	GEOVIDE	Pourquoi Pas	-12.219	40.333	200	5/23/2020	12.8188
nifH.G.13.035m.A	GEOVIDE	Pourquoi Pas	-13.888	41.384	35	5/25/2018	13.7655
nifH.G.15.020.B	GEOVIDE	Pourquoi Pas	-15.461	42	20	5/28/2016	14.9094
nifH.G.15.050m.C	GEOVIDE	Pourquoi Pas	-15.461	42	50	5/28/2018	14.8224
nifH.G.15.200m.A	GEOVIDE	Pourquoi Pas	-15.461	42	200	5/28/2020	12.1495
nifH.G.17.001m.A	GEOVIDE	Pourquoi Pas	-17.323	43.781	1	5/29/2014	14.6166
nifH.G.17.015m.A	GEOVIDE	Pourquoi Pas	-17.323	43.781	15	5/29/2016	14.4436
nifH.G.17.060m.B	GEOVIDE	Pourquoi Pas	-17.323	43.781	60	5/29/2018	12.5587
nifH.G.19.005m.C	GEOVIDE	Pourquoi Pas	-18.5	45.55	5	5/30/2014	15.0357
nifH.G.19.050m.A	GEOVIDE	Pourquoi Pas	-18.5	45.55	50	5/30/2018	12.8226
nifH.G.19.197m.B	GEOVIDE	Pourquoi Pas	-18.5	45.55	197	5/30/2020	12.0691
nifH.G.21.015m.A	GEOVIDE	Pourquoi Pas	-19.672	46.543	15	5/31/2015	14.0875
nifH.G.21.070m.A	GEOVIDE	Pourquoi Pas	-19.672	46.543	70	5/31/2017	12.6249
nifH.G.21.800m.C	GEOVIDE	Pourquoi Pas	-19.672	46.543	800	5/31/2019	8.8175
nifH.G.23.051m.A	GEOVIDE	Pourquoi Pas	-20.847	48.039	51	12/03/2014	13.241
nifH.G.23.200m.A	GEOVIDE	Pourquoi Pas	-20.847	48.039	200	14/03/2014	12.0954
nifH.G.25.056m.B	GEOVIDE	Pourquoi Pas	-22.172	49.529	56	12/03/2014	11.9825
nifH.G.25.200m.B	GEOVIDE	Pourquoi Pas	-22.172	49.529	200	14/03/2014	11.1109
nifH.G.26.007m.C	GEOVIDE	Pourquoi Pas	-22.602	50.278	7	06/04/2014	11.9248
nifH.G.26.500m.2B	GEOVIDE	Pourquoi Pas	-22.602	50.278	500	10/04/2014	6.5574
nifH.G.29.005m.C	GEOVIDE	Pourquoi Pas	-24.752	53.02	5	06/06/2014	10.1568
nifH.G.29.054m.A	GEOVIDE	Pourquoi Pas	-24.752	53.02	54	08/06/2014	8.4938
nifH.G.29.201m.B	GEOVIDE	Pourquoi Pas	-24.752	53.02	201	09/06/2014	7.0939
nifH.G.30.400m.A	GEOVIDE	Pourquoi Pas	-25.533	54	400	06/06/2014	5.2298
nifH.G.34.004m.B	GEOVIDE	Pourquoi Pas	-28.8786	57.004	4	06/09/2014	10.5137
nifH.G.34.020m.C	GEOVIDE	Pourquoi Pas	-28.8786	57.004	20	08/09/2014	10.1536
nifH.G.34.193m.B	GEOVIDE	Pourquoi Pas	-28.8786	57.004	193	11/09/2014	7.6817
nifH.G.36.003m.A	GEOVIDE	Pourquoi Pas	-29.7247	58.207	3	06/10/2014	9.5296
nifH.G.36.010m.B	GEOVIDE	Pourquoi Pas	-29.7247	58.207	10	07/10/2014	9.5281
nifH.G.36.050m.B	GEOVIDE	Pourquoi Pas	-29.7247	58.207	50	11/10/2014	7.8804
nifH.G.38.015m.C	GEOVIDE	Pourquoi Pas	-31.2665	58.843	15	08/10/2014	9.372
nifH.G.38.050m.A	GEOVIDE	Pourquoi Pas	-31.2665	58.843	50	10/10/2014	8.2349
nifH.G.38.650m.A	GEOVIDE	Pourquoi Pas	-31.2665	58.843	650	13/10/2014	6.0327
nifH.G.40.007m.A	GEOVIDE	Pourquoi Pas	-33.8291	59.102	7	06/12/2014	8.4263
nifH.G.40.030m.B	GEOVIDE	Pourquoi Pas	-33.8291	59.102	30	12/12/2014	7.7103
nifH.G.40.061m.B	GEOVIDE	Pourquoi Pas	-33.8291	59.102	61	19/12/2014	6.9511
nifH.G.42.030m.A	GEOVIDE	Pourquoi Pas	-36.3963	59.363	30	6/13/2020	6.8216
nifH.G.42.200m.A	GEOVIDE	Pourquoi Pas	-36.3963	59.363	200	6/13/2027	4.7656
nifH.G.44.070m.A	GEOVIDE	Pourquoi Pas	-38.954	59.623	70	6/14/2020	4.4351
nifH.G.53.005m.A	GEOVIDE	Pourquoi Pas	-43.0151	59.902	5	6/17/2014	-0.708
nifH.G.53.030m.B	GEOVIDE	Pourquoi Pas	-43.0151	59.902	30	6/17/2017	-0.8568

nifH.G.53.160m.A	GEOVIDE	Pourquoi Pas	-43.0151	59.902	160	6/17/2020	-0.6305
nifH.G.56.030m.A	GEOVIDE	Pourquoi Pas	-42.399	59.823	30	6/17/2017	4.0611
nifH.G.60.030m.B	GEOVIDE	Pourquoi Pas	-42.013	59.799	30	6/18/2016	6.6808
nifH.G.60.070m.B	GEOVIDE	Pourquoi Pas	-42.013	59.799	70	6/18/2017	6.2191
nifH.G.61.005m.A	GEOVIDE	Pourquoi Pas	-45.1122	59.753	5	6/18/2014	-0.0723
nifH.G.61.080m.A	GEOVIDE	Pourquoi Pas	-45.1122	59.753	80	6/18/2021	-1.0665
nifH.G.61.160m.A	GEOVIDE	Pourquoi Pas	-45.1122	59.753	160	6/18/2023	-0.0372
nifH.G.63.030m.A	GEOVIDE	Pourquoi Pas	-45.6891	59.434	30	6/19/2020	4.5185
nifH.G.63.100m.A	GEOVIDE	Pourquoi Pas	-45.6891	59.434	100	6/19/2023	4.7197
nifH.G.68.100m.A	GEOVIDE	Pourquoi Pas	-47.419	56.913	100	6/21/2023	3.8819
nifH.G.69.025m.C	GEOVIDE	Pourquoi Pas	-48.0935	55.841	25	6/22/2016	4.4109
nifH.G.69.1200m.A	GEOVIDE	Pourquoi Pas	-48.0935	55.841	1200	6/22/2018	3.4761
nifH.G.71.070m.A	GEOVIDE	Pourquoi Pas	-49.4333	53.692	70	6/24/2021	4.4551
nifH.G.71.200m.A	GEOVIDE	Pourquoi Pas	-49.4333	53.692	200	6/24/2025	3.7153
nifH.G.78.150m.A	GEOVIDE	Pourquoi Pas	-53.82	51.989	150	6/27/2021	-0.6399
nifH.lr3um.gDNA.002	M116	Meteor	-53.12211	12.36404	1	03/05/15	26.5
nifH.lr3um.gDNA.003	M116	Meteor	-52.21788	12.14585	1	04/05/15	26.4
nifH.lr3um.gDNA.004	M116	Meteor	-51.33403	11.53611	1	04/05/15	26
nifH.lr3um.gDNA.005	M116	Meteor	-50.45058	11.32632	1	04/05/15	26.2
nifH.lr3um.gDNA.006	M116	Meteor	-49.52415	11.08277	1	04/05/15	26
nifH.lr3um.gDNA.008	M116	Meteor	-48.1995	11	1	05/05/15	26.5
nifH.lr3um.gDNA.009	M116	Meteor	-47.37968	11	1	05/05/15	26.7
nifH.lr3um.gDNA.010	M116	Meteor	-46.41587	11	1	05/05/15	26.7
nifH.lr3um.gDNA.011	M116	Meteor	-46.1866	10.59937	1	06/05/15	25.6
nifH.lr3um.gDNA.012	M116	Meteor	-45.20009	11	1	06/05/15	25.4
nifH.lr3um.gDNA.013	M116	Meteor	-44.37125	11	1	06/05/15	25.7
nifH.lr3um.gDNA.014	M116	Meteor	-43.33661	10.55821	1	06/05/15	25.7
nifH.lr3um.gDNA.015	M116	Meteor	-42.25634	10.55732	1	07/05/15	26.3
nifH.lr3um.gDNA.016	M116	Meteor	-41.29372	11	1	07/05/15	26.3
nifH.lr3um.gDNA.017	M116	Meteor	-40.57927	10.33211	1	07/05/15	26.4
nifH.lr3um.gDNA.018	M116	Meteor	-40.53041	9.298	1	08/05/15	26.5
nifH.lr3um.gDNA.019	M116	Meteor	-40.47613	8.1915	1	08/05/15	26.7
nifH.lr3um.gDNA.020	M116	Meteor	-40.40202	7.5528	1	08/05/15	26.9
nifH.lr3um.gDNA.021	M116	Meteor	-40.18097	8.4265	1	08/05/15	26.6
nifH.lr3um.gDNA.022	M116	Meteor	-40.4175	9.3495	1	09/05/15	25.6
nifH.lr3um.gDNA.023	M116	Meteor	-40.0003	10.36219	1	09/05/15	25.8
nifH.lr3um.gDNA.024	M116	Meteor	-39.53994	10.56266	1	09/05/15	25.7
nifH.lr3um.gDNA.025	M116	Meteor	-39.25855	11.49013	1	10/05/15	25.1
nifH.lr3um.gDNA.026	M116	Meteor	-39.12666	11.38038	1	10/05/15	25.2
nifH.lr3um.gDNA.027	M116	Meteor	-37.33653	11	1	10/05/15	25
nifH.lr3um.gDNA.028	M116	Meteor	-36.22077	11	1	11/05/15	24.7
nifH.lr3um.gDNA.028.5	M116	Meteor	-36.22077	11	1	11/05/15	24.7
nifH.lr3um.gDNA.029	M116	Meteor	-35.47622	10.16879	1	11/05/15	25.7
nifH.lr3um.gDNA.030	M116	Meteor	-35.27683	9.07194	1	11/05/15	26.1
nifH.lr3um.gDNA.031	M116	Meteor	-35.12275	8.1314	1	12/05/15	26.5
nifH.lr3um.gDNA.032	M116	Meteor	-35	7.57552	1	12/05/15	26.6
nifH.lr3um.gDNA.033	M116	Meteor	-35	8.38487	1	12/05/15	26.2

nifH.lr3um.gDNA.034	M116	Meteor	-35	9.39643	1	12/05/15	26.1
nifH.lr3um.gDNA.035	M116	Meteor	-35	10.10521	1	13/05/15	25.9
nifH.lr3um.gDNA.036	M116	Meteor	-35	11.25137	1	13/05/15	25.6
nifH.lr3um.gDNA.037	M116	Meteor	-34.40295	11.40339	1	13/05/15	25.5
nifH.lr3um.gDNA.038	M116	Meteor	-33.35937	11	1	14/05/15/	25.3
nifH.lr3um.gDNA.039	M116	Meteor	-32.40528	11	1	14/05/15	24.7
nifH.lr3um.gDNA.040	M116	Meteor	-31.4472	11	1	14/05/15	25.7
nifH.lr3um.gDNA.041	M116	Meteor	-30.52724	11	1	15/05/15	25.3
nifH.lr3um.gDNA.042	M116	Meteor	-29.40018	11	1	15/05/15	25.3
nifH.lr3um.gDNA.043	M116	Meteor	-28.50489	10.12754	1	15/05/15	25.5
nifH.lr3um.gDNA.044	M116	Meteor	-28.39556	9.18265	1	15/05/15	26.4
nifH.lr3um.gDNA.045	M116	Meteor	-28.07187	6.36102	1	16/05/15	27.6
nifH.lr3um.gDNA.046	M116	Meteor	-28	6.15465	1	16/05/15	28
nifH.lr3um.gDNA.047	M116	Meteor	-28	7.04219	1	17/05/15	27.6
nifH.lr3um.gDNA.048	M116	Meteor	-28	8.16535	1	17/05/15	26.9
nifH.lr3um.gDNA.049	M116	Meteor	-28	8.48576	1	17/05/15	27
nifH.lr3um.gDNA.050	M116	Meteor	-28	9.08214	1	17/05/15	27.2
nifH.lr3um.gDNA.051	M116	Meteor	-28	9.52239	1	18/05/15	26.8
nifH.lr3um.gDNA.052	M116	Meteor	-28	10.35257	1	18/05/15	26
nifH.lr3um.gDNA.053	M116	Meteor	-28.00117	11.12392	1	18/05/15	25.7
nifH.lr3um.gDNA.054	M116	Meteor	-28.3295	11.59572	1	19/05/15	25.3
nifH.lr3um.gDNA.055	M116	Meteor	-28	13.48015	1	19/05/15	24.5
nifH.lr3um.gDNA.055.5	M116	Meteor	-28	13.48015	1	19/05/15	24.5
nifH.lr3um.gDNA.056	M116	Meteor	-27.39198	12.57874	1	19/05/15	25.5
nifH.lr3um.gDNA.057	M116	Meteor	-27.26265	12.1908	1	20/05/15	25.4
nifH.lr3um.gDNA.058	M116	Meteor	-26.38493	11	1	20/05/15	25.5
nifH.lr3um.gDNA.059	M116	Meteor	-26.0408	11	1	20/05/15	25.6
nifH.lr3um.gDNA.060	M116	Meteor	-25.15712	11	1	20/05/15	25.8
nifH.lr3um.gDNA.061	M116	Meteor	-24.48187	11	1	21/05/15	25.8
nifH.lr3um.gDNA.062	M116	Meteor	-23.42126	11.1405	1	21/05/15	25
nifH.lr3um.gDNA.063	M116	Meteor	-23.02409	11.27831	1	21/05/15	25
nifH.lr3um.gDNA.064	M116	Meteor	-23	11.41715	1	21/05/15	24.9
nifH.lr3um.gDNA.065	M116	Meteor	-23.00079	11.27702	1	22/05/15	24.8
nifH.lr3um.gDNA.066	M116	Meteor	-23	10.21029	1	22/05/15	26
nifH.lr3um.gDNA.067	M116	Meteor	-23	9.2638	1	22/05/15	26.5
nifH.lr3um.gDNA.068	M116	Meteor	-23	8.35512	1	22/05/15	26.2
nifH.lr3um.gDNA.069	M116	Meteor	-23	7.54035	1	23/05/15	27.2
nifH.lr3um.gDNA.070	M116	Meteor	-22	10.23768	1	24/05/15	26.7
nifH.lr3um.gDNA.071	M116	Meteor	-22.00111	11.04508	1	24/05/15	25.5
nifH.lr3um.gDNA.072	M116	Meteor	-22	11.29593	1	24/05/15	26.1
nifH.lr3um.gDNA.073	M116	Meteor	-21.42798	12	1	25/05/15	24.8
nifH.lr3um.gDNA.074	M116	Meteor	-21	11.07342	1	25/05/15	26.8
nifH.lr3um.gDNA.075	M116	Meteor	-21	10.4215	1	25/05/15	27.5
nifH.lr3um.gDNA.076	M116	Meteor	-21	9.32474	1	25/05/15	27.3
nifH.lr3um.gDNA.078	M116	Meteor	-21	7.57991	1	26/05/15	28.7
nifH.lr3um.gDNA.079	M116	Meteor	-21	7.06637	1	26/05/15	29.2
nifH.lr3um.gDNA.080	M116	Meteor	-21	6.11525	1	26/05/15	29

nifH.lr3um.gDNA.081	M116	Meteor	-21	5.08072	1	27/05/15	28.3
nifH.lr3um.gDNA.082	M116	Meteor	-20.42449	5.52736	1	27/05/15	28.9
nifH.lr3um.gDNA.083	M116	Meteor	-20.26928	6.39303	1	27/05/15	29.1
nifH.lr3um.gDNA.084	M116	Meteor	-20.05704	7.42921	1	27/05/15	28.9
nifH.lr3um.gDNA.085	M116	Meteor	-19.34481	8	1	28/05/15	29.1
nifH.lr3um.gDNA.086	M116	Meteor	-19	8.34789	1	28/05/15	29.1
nifH.lr3um.gDNA.087	M116	Meteor	-19.17557	9	1	28/05/15	29.1
nifH.lr3um.gDNA.088	M116	Meteor	-20	9.21345	1	28/05/15	28.3
nifH.lr3um.gDNA.089	M116	Meteor	-19.5643	10	1	29/05/15	27.8
nifH.lr3um.gDNA.090	M116	Meteor	-18.47329	10.12668	1	29/05/15	28
nifH.lr3um.gDNA.091	M116	Meteor	-18.0813	10.5186	1	29/05/15	28.4
nifH.lr3um.gDNA.092	M116	Meteor	-18.47	11	1	29/05/15	28.3
nifH.lr3um.gDNA.093	M116	Meteor	-19.07831	11	1	30/05/15	28.1
nifH.lr3um.gDNA.094	M116	Meteor	20	11.23954	1	30/05/15	26.3
nifH.lr3um.gDNA.095	M116	Meteor	-19.513	12	1	30/05/15	26.3
nifH.lr3um.gDNA.096	M116	Meteor	-19.05918	12	1	30/05/15	26.9
nifH.lr3um.gDNA.097	M116	Meteor	-19.35072	12.755	1	31/05/15	25.8
nifH.lr3um.gDNA.098	M116	Meteor	-20.4159	12.50805	1	31/05/15	24.9
nifH.lr3um.gDNA.099	M116	Meteor	-21	13.27118	1	31/05/15	25.1
nifH.lr3um.gDNA.100	M116	Meteor	-21	14.20423	1	31/05/15	24.6
nifH.lr3um.gDNA.101	M116	Meteor	-21.17454	15.13124	1	01/06/15	23.2
nifH.lr3um.gDNA.102	M116	Meteor	-21.51547	15.3873	1	01/06/15	22.8
nifH.lr3um.gDNA.103	M116	Meteor	-22.41482	16.19159	1	01/06/15	22.8
nifH.lr3um.gDNA.104	M116	Meteor	-23.16913	16.45296	1	01/06/15	23
nifH.lr3um.gDNA.105	M116	Meteor	-23.54763	17.16625	1	02/06/15	23
nifH.sm3um.gDNA.001	M116	Meteor	-53.52937	12.5402	1	03/05/15	26.8
nifH.sm3um.gDNA.002	M116	Meteor	-53.12211	12.36404	1	03/05/15	26.5
nifH.sm3um.gDNA.003	M116	Meteor	-52.21788	12.14585	1	04/05/15	26.4
nifH.sm3um.gDNA.004	M116	Meteor	-51.33403	11.53611	1	04/05/15	26
nifH.sm3um.gDNA.005	M116	Meteor	-50.45058	11.32632	1	04/05/15	26.2
nifH.sm3um.gDNA.006	M116	Meteor	-49.52415	11.08277	1	04/05/15	26
nifH.sm3um.gDNA.007	M116	Meteor	-49.15765	10.59999	1	05/05/15	26.3
nifH.sm3um.gDNA.008	M116	Meteor	-48.1995	11	1	05/05/15	26.5
nifH.sm3um.gDNA.009	M116	Meteor	-47.37968	11	1	05/05/15	26.7
nifH.sm3um.gDNA.010	M116	Meteor	-46.41587	11	1	05/05/15	26.7
nifH.sm3um.gDNA.011	M116	Meteor	-46.1866	10.59937	1	06/05/15	25.6
nifH.sm3um.gDNA.012	M116	Meteor	-45.20009	11	1	06/05/15	25.4
nifH.sm3um.gDNA.013	M116	Meteor	-44.37125	11	1	06/05/15	25.7
nifH.sm3um.gDNA.014	M116	Meteor	-43.33661	10.55821	1	06/05/15	25.7
nifH.sm3um.gDNA.015	M116	Meteor	-42.25634	10.55732	1	07/05/15	26.3
nifH.sm3um.gDNA.016	M116	Meteor	-41.29372	11	1	07/05/15	26.3
nifH.sm3um.gDNA.017	M116	Meteor	-40.57927	10.33211	1	07/05/15	26.4
nifH.sm3um.gDNA.018	M116	Meteor	-40.53041	9.298	1	08/05/15	26.5
nifH.sm3um.gDNA.019	M116	Meteor	-40.47613	8.1915	1	08/05/15	26.7
nifH.sm3um.gDNA.020	M116	Meteor	-40.40202	7.5528	1	08/05/15	26.9
nifH.sm3um.gDNA.021	M116	Meteor	-40.18097	8.4265	1	08/05/15	26.6
nifH.sm3um.gDNA.022	M116	Meteor	-40.4175	9.3495	1	09/05/15	25.6

nifH.sm3um.gDNA.023	M116	Meteor	-40.0003	10.36219	1	09/05/15	25.8
nifH.sm3um.gDNA.024	M116	Meteor	-39.53994	10.56266	1	09/05/15	25.7
nifH.sm3um.gDNA.025	M116	Meteor	-39.25855	11.49013	1	10/05/15	25.1
nifH.sm3um.gDNA.026	M116	Meteor	-39.12666	11.38038	1	10/05/15	25.2
nifH.sm3um.gDNA.027	M116	Meteor	-37.33653	11	1	10/05/15	25
nifH.sm3um.gDNA.028	M116	Meteor	-36.22077	11	1	11/05/15	24.7
nifH.sm3um.gDNA.029	M116	Meteor	-35.47622	10.16879	1	11/05/15	25.7
nifH.sm3um.gDNA.030	M116	Meteor	-35.27683	9.07194	1	11/05/15	26.1
nifH.sm3um.gDNA.031	M116	Meteor	-35.12275	8.1314	1	12/05/15	26.5
nifH.sm3um.gDNA.032	M116	Meteor	-35	7.57552	1	12/05/15	26.6
nifH.sm3um.gDNA.033	M116	Meteor	-35	8.38487	1	12/05/15	26.2
nifH.sm3um.gDNA.034	M116	Meteor	-35	9.39643	1	12/05/15	26.1
nifH.sm3um.gDNA.035	M116	Meteor	-35	10.10521	1	13/05/15	25.9
nifH.sm3um.gDNA.036	M116	Meteor	-35	11.25137	1	13/05/15	25.6
nifH.sm3um.gDNA.037	M116	Meteor	-34.40295	11.40339	1	13/05/15	25.5
nifH.sm3um.gDNA.038	M116	Meteor	-33.35937	11	1	14/05/15/	25.3
nifH.sm3um.gDNA.039	M116	Meteor	-32.40528	11	1	14/05/15	24.7
nifH.sm3um.gDNA.040	M116	Meteor	-31.4472	11	1	14/05/15	25.7
nifH.sm3um.gDNA.041	M116	Meteor	-30.52724	11	1	15/05/15	25.3
nifH.sm3um.gDNA.042	M116	Meteor	-29.40018	11	1	15/05/15	25.3
nifH.sm3um.gDNA.043	M116	Meteor	-28.50489	10.12754	1	15/05/15	25.5
nifH.sm3um.gDNA.044	M116	Meteor	-28.39556	9.18265	1	15/05/15	26.4
nifH.sm3um.gDNA.045	M116	Meteor	-28.07187	6.36102	1	16/05/15	27.6
nifH.sm3um.gDNA.046	M116	Meteor	-28	6.15465	1	16/05/15	28
nifH.sm3um.gDNA.047	M116	Meteor	-28	7.04219	1	17/05/15	27.6
nifH.sm3um.gDNA.048	M116	Meteor	-28	8.16535	1	17/05/15	26.9
nifH.sm3um.gDNA.049	M116	Meteor	-28	8.48576	1	17/05/15	27
nifH.sm3um.gDNA.050	M116	Meteor	-28	9.08214	1	17/05/15	27.2
nifH.sm3um.gDNA.051	M116	Meteor	-28	9.52239	1	18/05/15	26.8
nifH.sm3um.gDNA.052	M116	Meteor	-28	10.35257	1	18/05/15	26
nifH.sm3um.gDNA.053	M116	Meteor	-28.00117	11.12392	1	18/05/15	25.7
nifH.sm3um.gDNA.054	M116	Meteor	-28.3295	11.59572	1	19/05/15	25.3
nifH.sm3um.gDNA.055	M116	Meteor	-28	13.48015	1	19/05/15	24.5
nifH.sm3um.gDNA.056	M116	Meteor	-27.39198	12.57874	1	19/05/15	25.5
nifH.sm3um.gDNA.057	M116	Meteor	-27.26265	12.1908	1	20/05/15	25.4
nifH.sm3um.gDNA.058	M116	Meteor	-26.38493	11	1	20/05/15	25.5
nifH.sm3um.gDNA.059	M116	Meteor	-26.0408	11	1	20/05/15	25.6
nifH.sm3um.gDNA.060	M116	Meteor	-25.15712	11	1	20/05/15	25.8
nifH.sm3um.gDNA.061	M116	Meteor	-24.48187	11	1	21/05/15	25.8
nifH.sm3um.gDNA.062	M116	Meteor	-23.42126	11.1405	1	21/05/15	25
nifH.sm3um.gDNA.063	M116	Meteor	-23.02409	11.27831	1	21/05/15	25
nifH.sm3um.gDNA.064	M116	Meteor	-23	11.41715	1	21/05/15	24.9
nifH.sm3um.gDNA.065	M116	Meteor	-23.00079	11.27702	1	22/05/15	24.8
nifH.sm3um.gDNA.066	M116	Meteor	-23	10.21029	1	22/05/15	26
nifH.sm3um.gDNA.067	M116	Meteor	-23	9.2638	1	22/05/15	26.5
nifH.sm3um.gDNA.068	M116	Meteor	-23	8.35512	1	22/05/15	26.2
nifH.sm3um.gDNA.069	M116	Meteor	-23	7.54035	1	23/05/15	27.2

nifH.sm3um.gDNA.070	M116	Meteor	-22	10.23768	1	24/05/15	26.7
nifH.sm3um.gDNA.071	M116	Meteor	-22.00111	11.04508	1	24/05/15	25.5
nifH.sm3um.gDNA.072	M116	Meteor	-22	11.29593	1	24/05/15	26.1
nifH.sm3um.gDNA.073	M116	Meteor	-21.42798	12	1	25/05/15	24.8
nifH.sm3um.gDNA.074	M116	Meteor	-21	11.07342	1	25/05/15	26.8
nifH.sm3um.gDNA.075	M116	Meteor	-21	10.4215	1	25/05/15	27.5
nifH.sm3um.gDNA.076	M116	Meteor	-21	9.32474	1	25/05/15	27.3
nifH.sm3um.gDNA.077	M116	Meteor	-21	8.48313	1	26/05/15	28
nifH.sm3um.gDNA.078	M116	Meteor	-21	7.57991	1	26/05/15	28.7
nifH.sm3um.gDNA.079	M116	Meteor	-21	7.06637	1	26/05/15	29.2
nifH.sm3um.gDNA.080	M116	Meteor	-21	6.11525	1	26/05/15	29
nifH.sm3um.gDNA.081	M116	Meteor	-21	5.08072	1	27/05/15	28.3
nifH.sm3um.gDNA.082	M116	Meteor	-20.42449	5.52736	1	27/05/15	28.9
nifH.sm3um.gDNA.083	M116	Meteor	-20.26928	6.39303	1	27/05/15	29.1
nifH.sm3um.gDNA.084	M116	Meteor	-20.05704	7.42921	1	27/05/15	28.9
nifH.sm3um.gDNA.085	M116	Meteor	-19.34481	8	1	28/05/15	29.1
nifH.sm3um.gDNA.086	M116	Meteor	-19	8.34789	1	28/05/15	29.1
nifH.sm3um.gDNA.087	M116	Meteor	-19.17557	9	1	28/05/15	29.1
nifH.sm3um.gDNA.088	M116	Meteor	-20	9.21345	1	28/05/15	28.3
nifH.sm3um.gDNA.089	M116	Meteor	-19.5643	10	1	29/05/15	27.8
nifH.sm3um.gDNA.090	M116	Meteor	-18.47329	10.12668	1	29/05/15	28
nifH.sm3um.gDNA.091	M116	Meteor	-18.0813	10.5186	1	29/05/15	28.4
nifH.sm3um.gDNA.092	M116	Meteor	-18.47	11	1	29/05/15	28.3
nifH.sm3um.gDNA.093	M116	Meteor	-19.07831	11	1	30/05/15	28.1
nifH.sm3um.gDNA.094	M116	Meteor	20	11.23954	1	30/05/15	26.3
nifH.sm3um.gDNA.095	M116	Meteor	-19.513	12	1	30/05/15	26.3
nifH.sm3um.gDNA.096	M116	Meteor	-19.05918	12	1	30/05/15	26.9
nifH.sm3um.gDNA.097	M116	Meteor	-19.35072	12.755	1	31/05/15	25.8
nifH.sm3um.gDNA.098	M116	Meteor	-20.4159	12.50805	1	31/05/15	24.9
nifH.sm3um.gDNA.099	M116	Meteor	-21	13.27118	1	31/05/15	25.1
nifH.sm3um.gDNA.100	M116	Meteor	-21	14.20423	1	31/05/15	24.6
nifH.sm3um.gDNA.101	M116	Meteor	-21.17454	15.13124	1	01/06/15	23.2
nifH.sm3um.gDNA.102	M116	Meteor	-21.51547	15.3873	1	01/06/15	22.8
nifH.sm3um.gDNA.103	M116	Meteor	-22.41482	16.19159	1	01/06/15	22.8
nifH.sm3um.gDNA.104	M116	Meteor	-23.16913	16.45296	1	01/06/15	23
nifH.sm3um.gDNA.105	M116	Meteor	-23.54763	17.16625	1	02/06/15	23

Supplemental Table 4: Reference Genomes for the analysis of reference genomes in Figure 4.3

Reference Genome Number	Reference Genome	Taxonomy
1	Methanospirillum hungatei	Archaea
2	Methanoregula boonei	Archaea
3	Methanosphaerula palustris	Archaea
4	Methanosarcina barkeri	Archaea
5	Methanosarcina acetivorans	Archaea
6	Methanosarcina mazei	Archaea
7	Chloroherpeton thalassium	Chlorobi
8	Methanococcus maripaludis	Archaea
9	Acidithiobacillus ferrooxidans	Acidithiobacillia
10	Acidithiobacillus ferrooxidans	gamma-Proteobacteria
11	Dehalococcoides ethenogenes	Chloroflexi
12	methanogenic archaeon	Archaea
13	Methanothermobacter thermautotrophicus	Archaea
14	Methanococcus maripaludis	Archaea
15	Zymomonas mobilis	alpha-Proteobacteria
16	Zymomonas mobilis	alpha-Proteobacteria
17	Desulfovibrio africanus	delta-Proteobacteria
18	Pelobacter carbinolicus	delta-Proteobacteria
19	Clostridium kluyveri	Firmicutes
20	Methanosaeta concilii	Archaea
21	Desulfovibrio cf.	delta-Proteobacteria
22	Wolinella succinogenes	epsilon-Proteobacteria
23	Sulfuricurvum kujiense	epsilon-Proteobacteria
24	Geobacter uraniireducens	delta-Proteobacteria
25	Geobacter uraniireducens	delta-Proteobacteria
26	Desulfitobacterium hafniense	Firmicutes
27	Denitrovibrio acetiphilus	Deferribacteres
28	Pelodictyon phaeoclathratiforme	Chlorobi
29	Chlorobaculum parvum	Chlorobi
30	Chlorobium phaeobacteroides	Chlorobi
31	Prosthecochloris vibrioformis	Chlorobi
32	Chlorobium chlorochromatii	Chlorobi
33	Chlorobium tepidum	Chlorobi
34	Pelodictyon luteolum	Chlorobi
35	Chlorobium phaeobacteroides	Chlorobi
36	Desulfatibacillum alkenivorans	delta-Proteobacteria
37	Heliobacterium modesticaldum	Firmicutes
38	Geobacter bemidjiensis	delta-Proteobacteria
39	Desulfovibrio salexigens	delta-Proteobacteria
40	Desulfovibrio magneticus	delta-Proteobacteria
41	Desulfuromonas acetoxidans	delta-Proteobacteria
42	Desulfomicrobium baculatum	delta-Proteobacteria
43	Desulfovibrio vulgaris	delta-Proteobacteria
44	Desulfovibrio vulgaris	delta-Proteobacteria
45	Desulfotomaculum reducens	Firmicutes
46	Desulfovibrio vulgaris	delta-Proteobacteria
47	Desulfotomaculum acetoxidans	Firmicutes
48	Desulfotomaculum acetoxidans	Firmicutes
49	Desulforudis audaxviator	Firmicutes
50	Methanococcus maripaludis	Archaea
51	Halorhodospira halophila	gamma-Proteobacteria
52	Methylobacter tundripaludum	gamma-Proteobacteria
53	Aquabacterium sp.	beta-Proteobacteria
54	Accumulibacter phosphatis	beta-Proteobacteria
55	Geopsychrobacter electrodiphilus	delta-Proteobacteria
56	Dechlorosoma suillum	beta-Proteobacteria
57	Sedimenticola sp.	gamma-Proteobacteria
58	Beggiatoa alba	gamma-Proteobacteria
59	Allochromatium vinosum	gamma-Proteobacteria
60	Geobacter sp.	delta-Proteobacteria
61	Geobacter lovleyi	delta-Proteobacteria
62	Geobacter metallireducens	delta-Proteobacteria
63	Geobacter sulfurreducens	delta-Proteobacteria
64	Anaeromyxobacter sp.	delta-Proteobacteria

65	Anaeromyxobacter sp.	delta-Proteobacteria
66	Azoarcus sp.	beta-Proteobacteria
67	Methylobacter luteus	gamma-Proteobacteria
68	Methylococcus capsulatus	gamma-Proteobacteria
69	Pelobacter seleniigenes	delta-Proteobacteria
70	Novosphingobium sp.	alpha-Proteobacteria
71	Cyanothece sp.	Cyanobacteria
72	Cyanothece sp.	Cyanobacteria
73	Cyanothece sp.	Cyanobacteria
74	Cylindrospermum stagnale	Cyanobacteria
75	Calothrix sp.	Cyanobacteria
76	*Trichodesmium* erythraeum	Cyanobacteria
77	cyanobacterium endosymbiont	other
78	Cyanothece sp.	Cyanobacteria
79	Crocosphaera watsonii	Cyanobacteria
80	Leptolyngbya sp.	Cyanobacteria
81	Aphanocapsa montana	Cyanobacteria
82	Anabaena variabilis	Cyanobacteria
83	Nostoc sp.	Cyanobacteria
84	Cylindrospermopsis raciborskii	Cyanobacteria
85	Cyanothece sp.	Cyanobacteria
86	Nostoc punctiforme	Cyanobacteria
87	Rhodospirillum centenum	alpha-Proteobacteria
88	Methylacidiphilum infernorum	Verrucomicrobia
89	Synechococcus sp.	Cyanobacteria
90	Synechococcus sp.	Cyanobacteria
91	Methylocella silvestris	alpha-Proteobacteria
92	Rhodobacter sphaeroides	alpha-Proteobacteria
93	Rhodobacter sphaeroides	alpha-Proteobacteria
94	Rhodobacter sphaeroides	alpha-Proteobacteria
95	Leptothrix cholodnii	beta-Proteobacteria
96	Rhodomicrobium vannielii	alpha-Proteobacteria
97	Methyloversatilis thermotolerans	beta-Proteobacteria
98	Polaromonas naphthalenivorans	beta-Proteobacteria
99	Confluentimicrobium sp.	alpha-Proteobacteria
100	Rhodospirillum rubrum	alpha-Proteobacteria
101	Dechloromonas aromatica	beta-Proteobacteria
102	Clostridium acetobutylicum	Firmicutes
103	Thermoanaerobacterium thermosaccharolyticum	Firmicutes
104	Clostridium beijerinckii	Firmicutes
105	Alkaliphilus metalliredigens	Firmicutes
106	Caldicellulosiruptor saccharolyticus	Firmicutes
107	Roseiflexus castenholzi	Chloroflexi
108	Roseiflexus sp.	Chloroflexi
109	Frankia sp.	Actinobacteria
110	Bradyrhizobium sp.	alpha-Proteobacteria
111	Bradyrhizobium sp.	alpha-Proteobacteria
112	Bradyrhizobium japonicum	alpha-Proteobacteria
113	Rhodopseudomonas palustris	alpha-Proteobacteria
114	Rhodopseudomonas palustris	alpha-Proteobacteria
115	Rhodopseudomonas palustris	alpha-Proteobacteria
116	Xanthobacter sp.	alpha-Proteobacteria
117	Azorhizobium caulinodans	alpha-Proteobacteria
118	Rhizobium leguminosarum	alpha-Proteobacteria
119	Mesorhizobium loti	alpha-Proteobacteria
120	Sinorhizobium meliloti	alpha-Proteobacteria
121	Bradyrhizobium sp.	alpha-Proteobacteria
122	Pseudomonas stutzeri	gamma-Proteobacteria
123	Agarivorans gilvus	gamma-Proteobacteria
124	Teredinibacter turnerae	gamma-Proteobacteria
125	Martelella endophytica	alpha-Proteobacteria
126	Opitutaceae bacterium	Verrucomicrobia
127	Kosakonia sacchari	gamma-Proteobacteria
128	Dickeya dadantii	gamma-Proteobacteria
129	Erwinia carotovora	gamma-Proteobacteria
130	Burkholderia xenovorans	beta-Proteobacteria
131	Burkholderia vietnamiensis	beta-Proteobacteria
132	Azotobacter vinelandii	gamma-Proteobacteria

Supplemental Table 5: Metabolic Pathways for the analysis of reference genomes in Figure 4.3

Pathway Number	Metabolic Pathway
1	Pyruvate:ferredoxin oxidoreductase
2	Nitrogen fixation
3	Nitrate and nitrite ammonification
4	Methylglyoxal Metabolism
5	D-ribose utilization
6	Glycolate, glyoxylate interconversions
7	Polyphosphate
8	Mannose Metabolism
9	Glycogen metabolism
10	Pyruvate Alanine Serine Interconversions
11	TCA Cycle
12	Phosphate metabolism
13	Thioredoxin-disulfide reductase
14	Acetolactate synthase subunits
15	Quinate degradation
16	Anaerobic respiratory reductases
17	Glycolysis and Gluconeogenesis, including Archaeal enzymes
18	Pyruvate metabolism I: anaplerotic reactions, PEP
19	Ammonia assimilation
20	One-carbon metabolism by tetrahydropterines
21	Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate
22	High affinity phosphate transporter and control of PHO regulon
23	Pentose phosphate pathway
24	Carbon storage regulator
25	Fermentations: Mixed acid
26	Encapsulating protein for DyP-type peroxidase and ferritin-like protein oligomers
27	Acetoin, butanediol metabolism
28	Methanogenesis
29	Formaldehyde assimilation: Ribulose monophosphate pathway
30	Methanogenesis from methylated compounds
31	Sulfur oxidation
32	Sulfate reduction-associated complexes
33	Sulfite reduction-associated complex DsrMKJOP and co-clustering genes
34	Cyanobecterial bypass in the TCA
35	Trehalose Uptake and Utilization
36	Propionate-CoA to Succinate Module
37	Methylcitrate cycle
38	Iron siderophore sensor & receptor system
39	Dissimilatory nitrite reductase
40	Propionyl-CoA to Succinyl-CoA Module
41	Citrate Metabolism, Transport, and Regulation
42	Transport of Iron
43	Ferrous iron transporter EfeUOB, low-pH-induced
44	Soluble methane monooxygenase sMMO
45	Phenol hydroxylase
46	Toluene 4-monooxygenase TMO
47	Toluene degradation
48	Phosphoenolpyruvate phosphomutase
49	Taurine Utilization
50	D-galactonate catabolism
51	Nitrilase
52	L-fucose utilization temp
53	Allantoin Utilization
54	Siderophore Pyoverdine
55	2-Ketogluconate Utilization
56	Siderophore assembly kit
57	Sulfate assimilation related cluster
58	ABC transporter [iron.B.siderophore.hemin]
59	Melibiose Utilization
60	Erythritol utilization
61	Particulate methane monooxygenase pMMO
62	Energy conserving hydrogenase b, Methanococcales-Methanobacteriales-Methanopyrales
63	Archaeal membrane bound hydrogenases
64	Mebrane bound hydrogenases

65	Iron transport system including ABC transporter
66	Alpha-acetolactate operon
67	CitAB
68	L-Cystine Uptake and Metabolism
69	Siderophore Enterobactin
70	Dihydroxyacetone kinases
71	Alpha-Amylase locus in Streptococcoccus
72	Sucrose utilization
73	L-rhamnose utilization
74	Inositol catabolism
75	Mannitol Utilization
76	FructooligosaccharidesFOS and Raffinose Utilization
77	Beta-Glucoside Metabolism
78	L-Arabinose utilization
79	Biphenyl Degradation
80	VC0266
81	Acetone carboxylase
82	Ethylmalonyl-CoA pathway of C assimilation, GJO
83	Ethylmalonyl-CoA pathway of C assimilation
84	Siderophore Aerobactin
85	D-galactonate catabolism
86	Protocatechuate branch of beta-ketoadipate pathway
87	Chloroaromatic degradation pathway
88	Malonate decarboxylase
89	Central meta-cleavage pathway of aromatic compound degradation
90	D-galactarate, D-glucarate and D-glycerate catabolism
91	D-galactarate, D-glucarate and D-glycerate catabolism - gjo
92	Utilization of glutathione as a sulphur source
93	Homogentisate pathway of aromatic compound degradation
94	Tricarballylate Utilization
95	Benzoate transport and degradation cluster
96	Amidase clustered with urea and nitrile hydratase functions
97	4-Hydroxyphenylacetic acid catabolic pathway
98	Photorespiration oxidative C cycle
99	Calvin-Benson cycle
100	CO2 uptake, carboxysome
101	Galactosylceramide and Sulfatide metabolism
102	Lactose and Galactose Uptake and Utilization
103	Lactose utilization
104	Chitin and N-acetylglucosamine utilization
105	Entner-Doudoroff Pathway
106	Heme, hemin uptake and utilization systems in GramPositives
107	D-gluconate and ketogluconates metabolism
108	Deoxyribose and Deoxynucleoside Catabolism
109	Fermentations: Lactate
110	Trehalose Biosynthesis
111	Maltose and Maltodextrin Utilization
112	Glycerol and Glycerol-3-phosphate Uptake and Utilization
113	Serine-glyoxylate cycle
114	Nitrosative stress
115	Butanol Biosynthesis
116	Dehydrogenase complexes
117	Glycerate metabolism
118	Lactate utilization
119	Acetyl-CoA fermentation to Butyrate
120	Inorganic Sulfur Assimilation
121	Salicylate ester degradation
122	Salicylate and gentisate catabolism
123	Gentisate degradation
124	Glyoxylate bypass
125	Denitrification
126	Denitrifying reductase gene clusters
127	Cyanate hydrolysis
128	N-heterocyclic aromatic compound degradation
129	Aromatic Amin Catabolism
130	p-Hydroxybenzoate degradation
131	Catechol branch of beta-ketoadipate pathway
132	Iron acquisition in Streptococcus
133	Fructose utilization
134	D-Galacturonate and D-Glucuronate Utilization

135	Xylose utilization
136	Alkanesulfonates Utilization
137	Alkanesulfonate assimilation
138	Benzoate degradation
139	Alkylphosphonate utilization
140	Heme, hemin uptake and utilization systems in GramNegatives
141	Heme transport system

Supplemental Table 6: Pairwise statistical comparisons (ANOSIM) of taxonomic grouping with metabolic potential as defined using FROMP (Desai et al. 2013).

Group 1	Group 2	R statistic[1)]	Significance level (%)[1)]
alpha-Proteobacteria	Archaea	0.839	0.1
alpha-Proteobacteria	delta-Proteobacteria	0.442	0.1
Archaea	Chloroflexi	0.971	0.2
Archaea	delta-Proteobacteria	0.757	0.1
beta-Proteobacteria	Archaea	1	0.1
beta-Proteobacteria	Chloroflexi	0.826	0.3
beta-Proteobacteria	delta-Proteobacteria	0.395	0.1
Chlorobi	alpha-Proteobacteria	0.504	0.1
Chlorobi	Archaea	0.998	0.1
Chlorobi	beta-Proteobacteria	0.883	0.1
Cyanobacteria	alpha-Proteobacteria	0.467	0.1
Cyanobacteria	Archaea	0.869	0.1
Cyanobacteria	beta-Proteobacteria	0.74	0.1
Cyanobacteria	Chlorobi	0.738	0.1
Cyanobacteria	delta-Proteobacteria	0.576	0.1
Cyanobacteria	Firmicutes	0.693	0.1
Cyanobacteria	gamma-Proteobacteria	0.366	0.1
Firmicutes	alpha-Proteobacteria	0.563	0.1
Firmicutes	Archaea	0.99	0.1
Firmicutes	beta-Proteobacteria	0.711	0.1
Firmicutes	Chlorobi	0.445	0.1
Firmicutes	Chloroflexi	0.755	0.1
Firmicutes	delta-Proteobacteria	0.407	0.1
Firmicutes	gamma-Proteobacteria	0.4	0.1
gamma-Proteobacteria	alpha-Proteobacteria	0.244	0.1
gamma-Proteobacteria	Archaea	0.579	0.1
gamma-Proteobacteria	delta-Proteobacteria	0.332	0.1

1) An R value with a significance level lower than 5% indicates that groupings are significantly different from each other. R increases with significance.

Supplemental Table 7: Diazotrophic reference genomes with their assigned taxonomic group and carbon metabolism significantly associated with their taxonomic group.

Genome	Group	Ethylmalonyl-CoA pathway of C assimilation	Mannitol Utilization	Chitin and N-acetylglucosamine utilization	Glyoxylate bypass	Calvin-Benson cycle	Photorespiration oxidative C cycle	Lactate utilization	Deoxyribose and Deoxynucleoside Catabolism	Serine-glyoxylate cycle	Glycerate metabolism	D-gluconate and ketogluconates metabolism	Entner-Doudoroff Pathway	Maltose and Maltodextrin Utilization	Glycerol and Glycerol--phosphate Uptake and Utilization	D-ribose utilization	Carbon storage regulator	Acetyl-CoA fermentation to Butyrate	Fermentations: Mixed acid	Pyruvate: ferredoxin oxidoreductase	Fermentations: Lactate	Methylglyoxal Metabolism	Formaldehyde assimilation: Ribulose monophosphate	Methanogenesis	Trehalose Biosynthesis
Acidithiobacillus ferrooxidans ATCC 23270	Acidithiobacillia						+														+	+	+	+	
Frankia sp. EAN1pec	Actinobacteria	+	+	+				+	+			+	+	+	+			+			+				+
Azorhizobium caulinodans ORS 571	alpha-Proteobacteria	+					+	+				+	+	+	+			+	+			+			+
Bradyrhizobium japonicum USDA 110	alpha-Proteobacteria				+	+	+	+		+	+	+	+	+	+	+		+			+	+			+
Bradyrhizobium sp. BTAi1	alpha-Proteobacteria				+	+	+	+		+		+	+	+	+	+		+			+	+			+
Bradyrhizobium sp. ORS278	alpha-Proteobacteria				+	+	+	+				+	+	+	+	+		+	+		+	+	+		+
Bradyrhizobium sp. S23321	alpha-Proteobacteria		+		+	+	+	+		+	+	+	+	+	+	+					+	+			+
Confluentimicrobium sp. EMB200-NS6	alpha-Proteobacteria	+				+	+	+	+	+	+	+	+		+	+		+			+	+			+
Magnetococcus sp. MC-1	alpha-Proteobacteria									+				+			+	+		+		+			
Martelella endophytica	alpha-Proteobacteria		+	+	+			+	+	+	+	+	+	+	+	+		+	+		+	+			

Mesorhizobium loti MAFF303099	alpha-Proteobacteria		+	+	+			+	+	+	+	+	+		+	+		+				+	+			+
Methylocella silvestris BL2	alpha-Proteobacteria				+	+	+	+		+		+	+								+		+		+	+
Novosphingobium sp. MBES04	alpha-Proteobacteria							+			+	+			+			+					+			+
Rhizobium leguminosarum bv. viciae 3841	alpha-Proteobacteria		+	+	+			+	+	+	+	+	+	+	+	+		+				+	+			+
Rhodobacter sphaeroides 2.4.1	alpha-Proteobacteria	+		+		+	+	+	+	+	+	+	+		+	+		+				+	+			+
Rhodobacter sphaeroides ATCC 17025	alpha-Proteobacteria	+	+	+			+		+	+		+	+			+		+				+	+			+
Rhodobacter sphaeroides ATCC 17029	alpha-Proteobacteria	+	+	+		+	+	+	+	+			+			+		+				+	+			+
Rhodomicrobium vannielii ATCC 17100	alpha-Proteobacteria	+		+		+	+	+		+	+		+			+		+	+			+	+			+
Rhodopseudomonas palustris BisB5	alpha-Proteobacteria				+	+	+	+		+	+	+		+	+	+		+		+	+	+				+
Rhodopseudomonas palustris CGA009	alpha-Proteobacteria				+	+	+	+		+	+	+		+	+	+		+				+	+			+
Rhodopseudomonas palustris HaA2	alpha-Proteobacteria				+	+	+	+		+		+		+	+	+		+				+	+			+
Rhodospirillum centenum SW	alpha-Proteobacteria			+	+		+	+	+	+		+		+	+	+		+				+	+	+		+
Rhodospirillum rubrum	alpha-Proteobacteria	+				+	+	+	+	+	+			+	+	+		+	+				+			+
Sinorhizobium meliloti 1021	alpha-Proteobacteria		+	+	+	+	+	+	+	+	+	+	+	+	+	+		+				+	+			+
Xanthobacter sp. 126	alpha-Proteobacteria	+				+	+	+		+	+	+	+		+	+		+	+				+			+
Zymomonas mobilis subsp. mobilis ZM4	alpha-Proteobacteria							+				+	+						+			+	+			
Zymomonas mobilis subsp. pomaceae ATCC 29192	alpha-Proteobacteria			+				+		+	+	+	+						+			+	+			
Methanobacterium lacus	Archaea									+						+				+				+	+	
Methanococcus maripaludis C7	Archaea									+						+				+				+	+	
Methanococcus maripaludis S2	Archaea									+						+				+				+	+	
Methanoculleus marisnigri JR1	Archaea					+								+		+				+				+	+	+
methanogenic archaeon RC-I	Archaea								+							+				+				+	+	
Methanopyrus kandleri AV19	Archaea									+										+				+	+	
Methanoregula boonei 6A8	Archaea															+				+				+	+	
Methanosaeta concilii GP6	Archaea					+										+				+			+	+	+	
Methanosarcina acetivorans C2A	Archaea					+										+				+				+	+	

Methanosarcina barkeri str. Fusaro	Archaea					+										+				+			+	+	
Methanosarcina mazei Go1	Archaea					+										+				+			+	+	
Methanosphaerula palustris E1-9c	Archaea									+						+				+			+	+	
Methanospirillum hungatei JF-1	Archaea					+								+		+				+			+	+	
Methanothermobacter thermautotrophicus str. Delta H	Archaea								+	+						+			+	+			+	+	
Accumulibacter phosphatis clade IIA str. UW-1	beta-Proteobacteria	+			+	+	+	+		+				+	+	+		+	+	+		+			
Aquabacterium sp. NJ1	beta-Proteobacteria				+	+	+	+		+	+		+		+			+	+			+			+
Azoarcus sp. BH72	beta-Proteobacteria				+			+				+			+			+	+			+	+		+
Burkholderia vietnamiensis strain G4	beta-Proteobacteria			+	+			+		+		+	+	+	+	+		+			+	+			+
Burkholderia xenovorans LB400	beta-Proteobacteria			+	+	+	+	+	+	+	+	+	+	+	+	+		+			+	+			+
Dechloromonas aromatica RCB	beta-Proteobacteria				+		+	+		+	+							+	+	+	+	+			
Dechlorosoma suillum PS	beta-Proteobacteria				+			+		+	+					+		+	+		+	+			
Leptothrix cholodnii SP-6	beta-Proteobacteria			+	+	+	+	+				+	+		+			+		+		+			
Methyloversatilis thermotolerans 3t	beta-Proteobacteria				+		+	+		+	+	+				+		+	+		+	+			
Polaromonas naphthalenivorans CJ2	beta-Proteobacteria				+	+		+			+	+	+		+	+		+	+			+			+
Chlorobaculum parvum NCIB 8327	Chlorobi						+		+						+	+				+	+		+		+
Chlorobium chlorochromatii CaD3	Chlorobi						+		+							+				+				+	+
Chlorobium phaeobacteroides BS1	Chlorobi								+						+	+				+			+	+	+
Chlorobium phaeobacteroides DSM 266	Chlorobi						+		+						+	+				+				+	+
Chlorobium tepidum TLS	Chlorobi						+		+	+					+	+				+	+			+	+
Chloroherpeton thalassium ATCC 35110	Chlorobi								+	+						+				+	+		+	+	
Pelodictyon luteolum DSM 273	Chlorobi						+		+	+					+	+				+	+	+		+	+
Pelodictyon phaeoclathratiforme BU-1	Chlorobi								+	+					+	+				+	+	+	+	+	+
Prosthecochloris vibrioformis DSM 265	Chlorobi						+		+							+				+				+	+

Dehalococcoides ethenogenes 195	Chloroflexi														+				+		+		
Roseiflexus castenholzi DSM 13941	Chloroflexi	+		+		+	+	+			+		+		+		+		+	+	+		+
Roseiflexus sp. RS-1	Chloroflexi	+		+		+	+	+			+		+	+	+		+		+	+	+		+
Anabaena variabilis ATCC 29413	Cyanobacteria		+		+	+		+		+	+	+	+	+	+				+	+	+		+
Aphanocapsa montana BDHKU210001	Cyanobacteria		+			+		+		+	+		+		+			+	+	+	+		
Calothrix sp. 336/3	Cyanobacteria		+	+	+	+		+		+		+	+	+	+			+	+	+	+		
Crocosphaera watsonii WH 8501	Cyanobacteria		+		+	+	+	+	+	+	+	+	+		+			+	+	+	+		+
Cyanothece sp CCY 0110	Cyanobacteria		+		+	+		+	+		+	+	+	+	+				+	+	+		
Cyanothece sp PCC 7424	Cyanobacteria		+	+	+	+		+	+		+	+	+		+			+	+	+	+		+
Cyanothece sp PCC 7425	Cyanobacteria		+		+	+		+			+	+	+		+		+		+	+	+		+
Cyanothece sp PCC 8801	Cyanobacteria		+		+	+			+		+	+	+	+					+	+	+		+
Cyanothece sp. ATCC 51142	Cyanobacteria		+		+	+		+	+	+	+	+	+	+	+				+	+	+		+
Cylindrospermopsis raciborskii CS-505	Cyanobacteria		+		+	+						+	+		+			+	+		+		+
Cylindrospermum stagnale PCC 7417	Cyanobacteria		+		+	+		+	+	+	+	+	+	+	+				+	+	+		+
Leptolyngbya sp. PCC 7375	Cyanobacteria		+		+	+		+		+	+	+	+	+	+		+	+		+	+		+
Nostoc punctiforme PCC 73102	Cyanobacteria		+	+	+	+		+	+		+	+		+	+				+	+	+		+
Nostoc sp. PCC 7120	Cyanobacteria		+		+	+		+		+	+	+	+	+	+				+	+	+		+
Synechococcus sp. JA-2-3B'a(2-13)	Cyanobacteria				+			+				+	+		+			+		+		+	+
Synechococcus sp. JA-3-3Ab	Cyanobacteria				+			+				+	+		+			+		+	+	+	+
Trichodesmium erythraeum IMS101	Cyanobacteria		+		+	+	+	+		+	+	+	+	+	+						+		
Denitrovibrio acetiphilus DSM 12809	Deferribacteres								+	+					+	+		+	+		+		
Anaeromyxobacter sp. Fw109-5	delta-Proteobacteria			+			+	+			+		+	+			+		+	+	+		+
Anaeromyxobacter sp. K	delta-Proteobacteria			+				+	+				+	+	+		+		+		+	+	+
Desulfatibacillum alkenivorans AK-01	delta-Proteobacteria						+	+	+					+	+	+	+	+	+			+	
Desulfomicrobium baculatum DSM 4028	delta-Proteobacteria					+	+	+					+	+		+		+	+	+	+		

Desulfovibrio africanus str. Walvis Bay	delta-Proteobacteria						+	+		+			+		+	+			+	+	+			+
Desulfovibrio cf. magneticus IFRC170	delta-Proteobacteria						+												+					+
Desulfovibrio magneticus RS-1	delta-Proteobacteria						+						+	+	+	+		+	+		+		+	
Desulfovibrio salexigens DSM 2638	delta-Proteobacteria					+	+	+					+	+	+	+		+	+		+		+	
Desulfovibrio vulgaris str. 'Miyazaki F'	delta-Proteobacteria						+						+	+		+			+		+	+		
Desulfovibrio vulgaris subsp. vulgaris DP4	delta-Proteobacteria						+						+	+	+	+		+	+	+	+			
Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough	delta-Proteobacteria						+			+			+	+		+		+	+	+	+		+	
Desulfuromonas acetoxidans DSM 684	delta-Proteobacteria							+					+	+	+	+			+		+			
Geobacter bemidjiensis Bem	delta-Proteobacteria						+						+	+	+		+	+	+			+		+
Geobacter lovleyi SZ	delta-Proteobacteria						+			+			+	+	+						+			+
Geobacter metallireducens GS-15	delta-Proteobacteria								+				+	+	+	+	+	+	+					+
Geobacter sp. M18	delta-Proteobacteria						+		+	+			+	+	+		+	+			+			+
Geobacter sulfurreducens PCA	delta-Proteobacteria						+						+	+	+	+		+	+					+
Geobacter uraniireducens Rf4	delta-Proteobacteria						+	+	+	+			+		+	+	+		+		+			
Geobacter uraniireducens Rf4	delta-Proteobacteria						+						+		+	+	+		+				+	
Geopsychrobacter electrodiphilus DSM 16401	delta-Proteobacteria			+			+	+		+			+	+	+		+	+	+	+	+			
Pelobacter carbinolicus DSM 2380	delta-Proteobacteria							+		+			+			+		+			+			+
Pelobacter seleniigenes DSM 18267	delta-Proteobacteria						+			+	+		+	+		+	+	+		+	+			
Sulfuricurvum kujiense DSM 16994:	epsilon-Proteobacteria															+			+					
Wolinella succinogenes DSM 1740:	epsilon-Proteobacteria															+			+					
Alkaliphilus metalliredigens QYMF	Firmicutes	+	+					+				+		+	+	+	+	+	+	+	+			
Caldicellulosiruptor saccharolyticus DSM 8903	Firmicutes		+					+				+	+		+	+			+	+	+			+
Clostridium acetobutylicum ATCC 824	Firmicutes	+	+					+	+	+	+	+	+	+		+	+	+	+	+	+	+		+

Clostridium beijerinckii NCIMB 8052	Firmicutes	+	+				+	+				+	+	+	+	+	+	+	+	+		+		
Clostridium kluyveri DSM 555	Firmicutes							+					+			+	+	+	+	+	+			
Clostridium thermocellum ATCC 27405	Firmicutes		+					+		+						+			+	+	+			
Desulfitobacterium hafniense DCB-2	Firmicutes					+		+	+							+	+	+	+		+			
Desulforudis audaxviator MP104C	Firmicutes												+			+			+	+				+
Desulfotomaculum acetoxidans DSM 771	Firmicutes															+	+		+		+			
Desulfotomaculum acetoxidans DSM 771	Firmicutes															+	+		+		+			
Desulfotomaculum reducens MI-1	Firmicutes						+						+	+	+	+	+			+	+			
Heliobacterium modesticaldum Ice1	Firmicutes						+	+						+	+	+	+		+	+		+		
Thermoanaerobacterium thermosaccharolyticum DSM 571	Firmicutes		+					+		+	+	+	+	+	+	+		+	+	+	+	+		+
Acidithiobacillus ferrooxidans ATCC 53993	gamma-Proteobacteria				+	+														+	+	+	+	
Agarivorans gilvus	gamma-Proteobacteria	+	+	+				+	+	+	+	+	+	+	+	+	+	+		+	+			+
Allochromatium vinosum DSM 180	gamma-Proteobacteria			+	+	+							+		+	+		+	+	+	+			+
Azotobacter vinelandii	gamma-Proteobacteria			+			+		+	+	+	+	+	+	+	+	+			+				+
Beggiatoa alba B18LD	gamma-Proteobacteria			+	+	+	+			+			+		+	+	+	+			+			
Dickeya dadantii 3937	gamma-Proteobacteria	+	+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
Erwinia carotovora subsp. atroseptica SCRI1043	gamma-Proteobacteria	+	+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+
Halorhodospira halophila SL1	gamma-Proteobacteria			+	+	+							+	+		+	+		+		+			
Kosakonia sacchari SP1	gamma-Proteobacteria	+	+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+
Methylobacter luteus IMV-B-3098	gamma-Proteobacteria								+	+	+	+	+	+	+	+	+		+	+	+	+		
Methylobacter tundripaludum SV96	gamma-Proteobacteria													+	+		+		+					
Methylococcus capsulatus str. Bath	gamma-Proteobacteria				+	+			+	+		+	+	+	+	+			+	+	+	+		

Pseudomonas stutzeri A1501	gamma-Proteobacteria		+			+		+	+	+	+	+	+		+	+			+	+			+
Sedimenticola sp. SIP-G1	gamma-Proteobacteria		+			+		+	+				+	+	+	+	+	+		+			
Teredinibacter turnerae T7901	gamma-Proteobacteria	+	+			+		+	+	+	+	+	+	+	+	+	+	+	+	+			+
cyanobacterium endosymbiont of Epithemia turgida isolate EtSB Lake Yunoko	Cyanobacteria	+		+						+	+	+	+	+					+	+			+
Methylacidiphilum infernorum V4	Verrucomicrobia	+			+		+					+	+	+				+	+	+	+		+
Opitutaceae bacterium TAV5	Verrucomicrobia	+				+			+	+	+	+	+	+	+		+	+	+	+			+

Supplemental Table 8: Diazotrophic reference genomes with their assigned taxonomic group and aromatic compound, nitrogen, phosphor, sulfur and iron related metabolism significantly associated with their taxonomic group.

Genome	Group	Benzoate degradation	N-heterocyclic aromatic compound degradation	Gentisate degradation	Salicylate ester degradation	Denitrifying reductase gene clusters	Nitrate and nitrite ammonification	Alkylphosphonate utilization	Utilization of glutathione as a sulphur source	Inorganic Sulfur Assimilation	Sulfite reduction-associated complex DsrMKJOP and co-clustering genes	Salmochelin-mediated Iron Acquisition	Siderophore assembly kit	Ferrous iron transporter EfeUOB, low-pH-induced	Heme, hemin uptake and utilization systems in GramPositives	Hemin transport system
Acidithiobacillus ferrooxidans ATCC 23270	Acidithiobacillia							+		+						
Frankia sp. EAN1pec	Actinobacteria			+	+		+		+	+		+	+	+		
Azorhizobium caulinodans ORS 571	alpha-Proteobacteria		+	+	+	+	+	+	+	+					+	+
Bradyrhizobium japonicum USDA 110	alpha-Proteobacteria		+	+		+	+	+	+	+					+	+
Bradyrhizobium sp. BTAi1	alpha-Proteobacteria	+	+	+	+	+	+	+	+	+					+	
Bradyrhizobium sp. ORS278	alpha-Proteobacteria	+	+	+	+	+	+	+	+	+					+	+
Bradyrhizobium sp. S23321	alpha-Proteobacteria	+	+	+	+	+	+	+							+	+

Confluentimicrobium sp. EMB200-NS6	alpha-Proteobacteria	+		+	+			+					+			+
Magnetococcus sp. MC-1	alpha-Proteobacteria					+					+					
Martelella endophytica	alpha-Proteobacteria	+		+	+	+	+	+	+	+				+		+
Mesorhizobium loti MAFF303099	alpha-Proteobacteria	+	+	+	+		+	+		+					+	+
Methylocella silvestris BL2	alpha-Proteobacteria		+		+	+	+			+				+	+	+
Novosphingobium sp. MBES04	alpha-Proteobacteria	+	+	+	+		+									
Rhizobium leguminosarum bv. viciae 3841	alpha-Proteobacteria	+	+	+	+		+	+	+						+	+
Rhodobacter sphaeroides 2.4.1	alpha-Proteobacteria	+	+			+		+							+	+
Rhodobacter sphaeroides ATCC 17025	alpha-Proteobacteria	+	+	+	+	+									+	+
Rhodobacter sphaeroides ATCC 17029	alpha-Proteobacteria	+	+	+	+	+		+	+						+	+
Rhodomicrobium vannielii ATCC 17100	alpha-Proteobacteria	+			+			+		+	+				+	+
Rhodopseudomonas palustris BisB5	alpha-Proteobacteria	+		+	+		+	+	+	+					+	+
Rhodopseudomonas palustris CGA009	alpha-Proteobacteria	+		+	+	+	+	+	+	+			+		+	+
Rhodopseudomonas palustris HaA2	alpha-Proteobacteria		+	+	+	+		+	+						+	
Rhodospirillum centenum SW	alpha-Proteobacteria					+				+					+	+
Rhodospirillum rubrum	alpha-Proteobacteria	+		+				+		+				+		
Sinorhizobium meliloti 1021	alpha-Proteobacteria	+	+	+	+	+	+	+		+			+		+	+
Xanthobacter sp. 126	alpha-Proteobacteria	+	+	+	+	+	+	+	+	+					+	+
Zymomonas mobilis subsp. mobilis ZM4	alpha-Proteobacteria							+		+						
Zymomonas mobilis subsp. pomaceae ATCC 29192	alpha-Proteobacteria	+														
Methanobacterium lacus	Archaea			+	+											
Methanococcus maripaludis C7	Archaea															
Methanococcus maripaludis S2	Archaea															
Methanoculleus marisnigri JR1	Archaea															
methanogenic archaeon RC-I	Archaea															
Methanopyrus kandleri AV19	Archaea															
Methanoregula boonei 6A8	Archaea															
Methanosaeta concilii GP6	Archaea			+			+									

Methanosarcina acetivorans C2A	Archaea															
Methanosarcina barkeri str. Fusaro	Archaea															
Methanosarcina mazei Go1	Archaea															
Methanosphaerula palustris E1-9c	Archaea															
Methanospirillum hungatei JF-1	Archaea															
Methanothermobacter thermautotrophicus str. Delta H	Archaea															
Accumulibacter phosphatis clade IIA str. UW-1	beta-Proteobacteria	+	+	+		+	+			+						+
Aquabacterium sp. NJ1	beta-Proteobacteria	+	+	+	+	+	+	+		+						+
Azoarcus sp. BH72	beta-Proteobacteria	+	+	+		+	+			+						
Burkholderia vietnamiensis strain G4	beta-Proteobacteria	+	+	+		+	+	+		+						+
Burkholderia xenovorans LB400	beta-Proteobacteria	+	+	+	+		+	+		+					+	+
Dechloromonas aromatica RCB	beta-Proteobacteria	+	+			+	+									
Dechlorosoma suillum PS	beta-Proteobacteria	+		+		+	+			+						
Leptothrix cholodnii SP-6	beta-Proteobacteria		+	+	+	+	+	+								+
Methyloversatilis thermotolerans 3t	beta-Proteobacteria		+	+			+	+		+					+	+
Polaromonas naphthalenivorans CJ2	beta-Proteobacteria	+		+	+	+	+	+		+						+
Chlorobaculum parvum NCIB 8327	Chlorobi										+					
Chlorobium chlorochromatii CaD3	Chlorobi										+					
Chlorobium phaeobacteroides BS1	Chlorobi										+					
Chlorobium phaeobacteroides DSM 266	Chlorobi															
Chlorobium tepidum TLS	Chlorobi										+					
Chloroherpeton thalassium ATCC 35110	Chlorobi															
Pelodictyon luteolum DSM 273	Chlorobi										+					
Pelodictyon phaeoclathratiforme BU-1	Chlorobi										+					
Prosthecochloris vibrioformis DSM 265	Chlorobi										+					
Dehalococcoides ethenogenes 195	Chloroflexi						+									
Roseiflexus castenholzi DSM 13941	Chloroflexi														+	
Roseiflexus sp. RS-1	Chloroflexi			+											+	

Anabaena variabilis ATCC 29413	Cyanobacteria				+		+			+				+			+	+	
Aphanocapsa montana BDHKU210001	Cyanobacteria			+	+		+	+		+							+		
Calothrix sp. 336/3	Cyanobacteria			+	+		+			+							+		
Crocosphaera watsonii WH 8501	Cyanobacteria				+		+									+	+	+	
Cyanothece sp CCY 0110	Cyanobacteria						+			+							+		
Cyanothece sp PCC 7424	Cyanobacteria			+			+			+							+		
Cyanothece sp PCC 7425	Cyanobacteria			+			+			+							+		
Cyanothece sp PCC 8801	Cyanobacteria						+			+							+		
Cyanothece sp. ATCC 51142	Cyanobacteria				+		+			+							+	+	
Cylindrospermopsis raciborskii CS-505	Cyanobacteria			+			+	+									+		
Cylindrospermum stagnale PCC 7417	Cyanobacteria			+	+		+			+							+		
Leptolyngbya sp. PCC 7375	Cyanobacteria	+			+		+	+		+							+		
Nostoc punctiforme PCC 73102	Cyanobacteria						+										+		
Nostoc sp. PCC 7120	Cyanobacteria						+	+						+			+		
Synechococcus sp. JA-2-3B'a(2-13)	Cyanobacteria						+	+		+							+	+	
Synechococcus sp. JA-3-3Ab	Cyanobacteria						+			+							+	+	
Trichodesmium erythraeum IMS101	Cyanobacteria				+		+	+		+							+		
Denitrovibrio acetiphilus DSM 12809	Deferribacteres			+		+	+											+	
Anaeromyxobacter sp. Fw109-5	delta-Proteobacteria				+	+				+									
Anaeromyxobacter sp. K	delta-Proteobacteria		+	+	+	+	+											+	
Desulfatibacillum alkenivorans AK-01	delta-Proteobacteria										+								
Desulfomicrobium baculatum DSM 4028	delta-Proteobacteria										+								
Desulfovibrio africanus str. Walvis Bay	delta-Proteobacteria			+	+		+				+								
Desulfovibrio cf. magneticus IFRC170	delta-Proteobacteria				+		+												
Desulfovibrio magneticus RS-1	delta-Proteobacteria			+			+												
Desulfovibrio salexigens DSM 2638	delta-Proteobacteria			+			+											+	
Desulfovibrio vulgaris str. 'Miyazaki F'	delta-Proteobacteria										+								
Desulfovibrio vulgaris subsp. vulgaris DP4	delta-Proteobacteria										+								

Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough	delta-Proteobacteria						+				+					
Desulfuromonas acetoxidans DSM 684	delta-Proteobacteria			+			+									+
Geobacter bemidjiensis Bem	delta-Proteobacteria					+										
Geobacter lovleyi SZ	delta-Proteobacteria			+		+	+									+
Geobacter metallireducens GS-15	delta-Proteobacteria		+			+	+									
Geobacter sp. M18	delta-Proteobacteria			+	+	+	+			+				+		
Geobacter sulfurreducens PCA	delta-Proteobacteria		+				+			+						
Geobacter uraniireducens Rf4	delta-Proteobacteria						+		+	+						+
Geobacter uraniireducens Rf4	delta-Proteobacteria						+		+	+						
Geopsychrobacter electrodiphilus DSM 16401	delta-Proteobacteria			+		+	+			+				+		+
Pelobacter carbinolicus DSM 2380	delta-Proteobacteria			+	+		+			+						+
Pelobacter seleniigenes DSM 18267	delta-Proteobacteria			+	+	+	+									+
Sulfuricurvum kujiense DSM 16994:	epsilon-Proteobacteria				+		+			+						
Wolinella succinogenes DSM 1740:	epsilon-Proteobacteria					+	+			+						
Alkaliphilus metalliredigens QYMF	Firmicutes					+		+		+						
Caldicellulosiruptor saccharolyticus DSM 8903	Firmicutes									+						
Clostridium acetobutylicum ATCC 824	Firmicutes															
Clostridium beijerinckii NCIMB 8052	Firmicutes															
Clostridium kluyveri DSM 555	Firmicutes						+		+							
Clostridium thermocellum ATCC 27405	Firmicutes						+									
Desulfitobacterium hafniense DCB-2	Firmicutes	+				+	+	+	+		+					+
Desulforudis audaxviator MP104C	Firmicutes										+					
Desulfotomaculum acetoxidans DSM 771	Firmicutes										+					
Desulfotomaculum acetoxidans DSM 771	Firmicutes			+			+				+					
Desulfotomaculum reducens MI-1	Firmicutes										+					
Heliobacterium modesticaldum Ice1	Firmicutes						+									
Thermoanaerobacterium thermosaccharolyticum DSM 571	Firmicutes						+									
Acidithiobacillus ferrooxidans ATCC 53993	gamma-Proteobacteria							+		+						

Agarivorans gilvus	gamma-Proteobacteria	+		+			+	+			+					+
Allochromatium vinosum DSM 180	gamma-Proteobacteria		+		+						+					+
Azotobacter vinelandii	gamma-Proteobacteria	+	+	+			+	+		+	+		+			+
Beggiatoa alba B18LD	gamma-Proteobacteria		+	+	+		+			+	+					+
Dickeya dadantii 3937	gamma-Proteobacteria			+		+	+	+	+	+	+			+	+	+
Erwinia carotovora subsp. atroseptica SCRI1043	gamma-Proteobacteria	+				+	+	+		+	+				+	+
Halorhodospira halophila SL1	gamma-Proteobacteria										+					
Kosakonia sacchari SP1	gamma-Proteobacteria	+		+		+	+	+	+	+	+			+	+	+
Methylobacter luteus IMV-B-3098	gamma-Proteobacteria		+	+		+	+	+		+	+					+
Methylobacter tundripaludum SV96	gamma-Proteobacteria		+			+	+	+								
Methylococcus capsulatus str. Bath	gamma-Proteobacteria					+	+			+	+					
Pseudomonas stutzeri A1501	gamma-Proteobacteria	+	+	+		+	+	+		+	+		+			+
Sedimenticola sp. SIP-G1	gamma-Proteobacteria			+	+	+	+	+			+					
Teredinibacter turnerae T7901	gamma-Proteobacteria	+					+	+		+						+
cyanobacterium endosymbiont of Epithemia turgida isolate EtSB Lake Yunoko	Cyanobacteria			+	+					+						
Methylacidiphilum infernorum V4	Verrucomicrobia						+			+						
Opitutaceae bacterium TAV5	Verrucomicrobia			+		+	+			+						+

Supplemental Table 9: Results of CD-HIT cluster analysis.

Cluster		GI numbers occurring in cluster	Clusters name as appears in tree (*) or sequence similarity
Cluster 1			
	1	EF204559	*
Cluster 2			
	1	273121.1	*
Cluster 3			
	1	HQ456000	*
Cluster 4			
	1	HQ455865	*
	2	HQ455837	97.80%
Cluster 5			
	1	HQ456094	*
Cluster 6			
	1	HQ660939	*
Cluster 7			
	1	EF568535	*
	2	EF568533	99.72%
Cluster 8			
	1	709032.9	*
Cluster 9			
	1	8774551	*
Cluster 10			
	1	439235b3	*
Cluster 11			
	1	AY896353	*
	2	KF151468	96.30%
	3	JQ358707	99.38%
	4	JF429940	99.07%
	5	DQ481332	96.91%
	6	DQ481321	97.22%
	7	DQ481317	97.53%
	8	DQ481316	96.60%
	9	DQ481315	96.60%
	10	KC013225	98.46%
	11	KC013169	99.07%
	12	KC013149	99.07%
	13	EU052680	99.07%
	14	EU052679	98.77%
	15	EU052677	99.07%
	16	EU052676	99.07%
	17	EU052672	99.07%
	18	EU052671	99.07%
	19	EU052670	98.46%
	20	EU052669	98.77%
	21	EU052667	98.77%
	22	EU052665	99.07%
	23	EU052663	98.50%
	24	EU052662	99.07%
	25	EU052596	98.77%
	26	EU052595	98.77%
	27	EU052593	98.77%
	28	EU052589	99.07%
	29	EU052588	99.07%
	30	EU052587	98.46%
	31	EU052586	98.77%
	32	EU052582	99.07%
	33	EU052581	98.46%
	34	EU052580	99.07%
	35	EU052579	99.07%
	36	EU052574	98.77%
	37	EU052573	99.07%
	38	EU052572	98.77%
	39	EU052567	99.07%
	40	EU052566	99.07%
	41	EU052565	99.07%
	42	EU052560	98.46%
	43	EU052559	98.77%

44	EU052557	98.46%
45	EU052556	98.77%
46	EU052554	98.77%
47	EU052553	98.77%
48	EU052547	99.07%
49	EU052546	99.07%
50	EU052545	98.77%
51	EU052542	98.77%
52	EU052541	99.07%
53	EU052540	99.07%
54	EU052539	98.46%
55	EU052536	98.77%
56	EU052535	98.46%
57	EU052534	98.46%
58	EU052533	98.77%
59	EU052528	98.46%
60	EU052527	99.07%
61	EU052525	99.07%
62	EU052436	99.07%
63	EU052435	98.46%
64	EU052434	99.07%
65	EU052432	98.50%
66	EU052431	98.46%
67	EU052430	98.77%
68	EU052429	99.07%
69	EU052427	99.07%
70	EU052426	99.07%
71	EU052424	99.07%
72	EU052422	99.07%
73	EU052421	99.07%
74	EU052410	99.07%
75	EU052407	98.77%
76	EU052401	98.77%
77	EU052400	99.07%
78	EU052398	98.77%
79	EU052397	99.07%
80	EU052378	99.07%
81	EU052377	99.07%
82	EU052376	98.46%
83	EU052375	99.07%
84	EU052374	99.07%
85	EU052372	98.50%
86	EU052371	99.07%
87	EU052370	99.07%
88	EU052369	99.07%
89	EU052368	99.07%
90	EU052367	98.77%
91	EU052366	98.77%
92	EU052364	99.07%
93	EU052362	99.07%
94	EU052361	99.07%
95	EU052360	98.46%
96	EU052358	98.50%
97	EU052356	98.77%
98	EU052353	99.07%
99	EU052352	99.07%
100	EU052349	98.46%
101	EU052348	99.38%
102	EU052340	99.07%
103	EU052339	99.07%
104	EU052337	99.07%
105	EU052334	98.77%
106	EU052332	98.77%
107	EU052331	98.77%
108	EU052328	99.07%
109	EU052315	99.07%
110	EU052314	99.07%
111	EU052312	98.77%
112	EU052309	99.07%
113	EU052307	97.53%

114	EU052305	98.77%
115	EU052303	98.77%
116	EU052302	98.77%
117	EU052301	98.77%
118	EU052300	98.46%
119	EU052299	98.77%
120	EU052298	98.77%
121	AB928257	99.07%
122	AB928304	96.60%
123	DQ825744	98.77%
124	DQ825740	98.77%
125	KF151469	96.60%
126	203124	96.91%
127	HQ611956	96.60%
128	HQ611934	96.60%
129	HQ611924	96.91%
130	HQ611922	96.60%
131	HQ611917	96.60%
132	HQ611906	96.91%
133	HQ611889	96.91%
134	HQ611881	96.91%
135	HQ611877	96.91%
136	HQ611871	96.91%
137	HQ611868	96.91%
138	HQ611859	96.91%
139	HQ611858	96.60%
140	HQ611854	96.91%
141	HQ611852	96.91%
142	HQ611847	96.91%
143	HQ611842	96.60%
144	HQ611809	97.53%
145	HQ611803	96.91%
146	HQ611771	97.22%
147	HQ611763	97.53%
148	HQ611759	96.91%
149	HQ611752	96.88%
150	HQ611746	97.53%
151	HQ611571	97.53%
152	HQ611568	97.22%
153	HQ611566	97.53%
154	HQ611563	96.91%
155	HQ611561	99.07%
156	HQ611558	97.22%
157	HQ611555	97.22%
158	HQ611550	99.07%
159	HQ611545	98.77%
160	HQ611544	97.84%
161	HQ611541	99.38%
162	HQ611540	97.53%
163	HQ611536	97.53%
164	HQ611527	98.77%
165	HQ611524	99.07%
166	HQ611523	99.07%
167	HQ611521	99.07%
168	HQ611520	98.77%
169	HQ611519	98.77%
170	HQ611514	98.50%
171	HQ611513	99.07%
172	HQ611511	97.22%
173	HQ611508	96.91%
174	HQ611491	99.38%
175	HQ611486	97.22%
176	HQ611483	97.22%
177	HQ611479	98.77%
178	HQ611471	98.50%
179	HQ611466	97.22%
180	HQ611465	99.07%
181	HQ611463	98.46%
182	HQ611462	99.38%
183	HQ611461	99.07%

184	HQ611451	99.38%
185	HQ611450	99.07%
186	HQ611445	97.22%
187	HQ611443	97.53%
188	HQ611441	97.22%
189	HQ611428	97.22%
190	HQ611417	98.77%
191	HQ611410	99.07%
192	HQ611402	97.53%
193	HQ611399	96.60%
194	HQ611397	97.53%
195	HQ611394	97.53%
196	HQ611391	96.91%
197	HQ456046	99.07%
198	HQ456036	98.77%
199	HQ456035	98.77%
200	HQ456015	96.91%
201	HQ456014	99.07%
202	HQ456010	99.07%
203	HQ456008	97.53%
204	HQ456005	98.77%
205	HQ455992	99.38%
206	HQ455988	99.69%
207	HQ455976	97.53%
208	HQ455965	98.50%
209	HQ455963	97.22%
210	HQ455959	99.38%
211	HQ455957	99.07%
212	HQ455953	96.91%
213	HQ455947	97.22%
214	HQ455924	97.22%
215	HQ455915	99.38%
216	HQ455898	99.38%
217	HQ455895	96.30%
218	HQ455887	98.77%
219	HQ455880	99.38%
220	GQ475484	98.77%
221	GQ475483	99.07%
222	GQ475482	98.77%
223	GQ475481	99.39%
224	GQ475480	99.07%
225	GQ475467	99.07%
226	GQ475446	98.77%
227	GQ475445	99.07%
228	GQ475435	98.77%
229	GQ475434	98.77%
230	GQ475433	99.07%
231	GQ475432	98.77%
232	GQ475431	99.07%
233	AY821848	99.38%
234	AY821847	99.07%
235	AY896460	97.22%
236	AY896459	96.63%
237	AY896458	96.62%
238	AY896457	96.93%
239	AY896453	97.22%
240	AY896452	97.22%
241	AY896451	97.22%
242	AY896450	97.22%
243	AY896449	96.91%
244	AY896448	97.22%
245	AY896447	97.22%
246	AY896446	97.22%
247	AY896445	99.38%
248	AY896444	97.22%
249	AY896443	99.07%
250	AY896442	99.38%
251	AY896441	98.46%
252	AY896416	97.24%
253	AY896415	97.24%

254	AY896413	96.93%
255	AY896412	97.24%
256	AY896411	96.63%
257	AY896409	97.55%
258	AY896408	96.01%
259	AY896407	98.77%
260	AY896405	97.21%
261	AY896404	99.08%
262	AY896403	97.29%
263	AY896402	99.08%
264	AY896401	99.08%
265	AY896400	99.08%
266	AY896399	97.85%
267	AY896398	97.55%
268	AY896397	98.77%
269	AY896396	98.77%
270	AY896395	99.39%
271	AY896394	99.07%
272	AY896393	98.50%
273	AY896392	96.91%
274	AY896391	98.77%
275	AY896390	99.07%
276	AY896389	98.50%
277	AY896388	96.91%
278	AY896387	96.93%
279	AY896386	97.85%
280	AY896385	96.93%
281	AY896383	99.08%
282	AY896382	99.08%
283	AY896381	96.93%
284	AY896380	97.24%
285	AY896379	97.24%
286	AY896378	99.08%
287	AY896376	98.77%
288	AY896375	96.93%
289	AY896374	96.93%
290	AY896373	96.93%
291	AY896368	96.93%
292	AY896367	97.55%
293	AY896366	96.32%
294	AY896364	97.24%
295	AY896363	99.08%
296	AY896362	99.08%
297	AY896352	97.24%
298	AY896350	96.63%
299	AY896349	98.50%
300	AY896348	99.08%
301	AY896347	96.93%
302	AY896346	96.90%
303	AY896345	97.89%
304	AY896344	97.23%
305	AY896339	98.77%
306	AY896338	97.24%
307	AY896337	96.99%
308	AY896335	97.55%
309	AY896333	98.77%
310	AY896332	97.85%
311	AY896331	98.47%
312	AY896328	99.08%
313	AY896327	99.08%
314	AY896326	97.24%
315	AY896325	97.24%
316	AY896324	98.47%
317	AY896323	97.24%
318	AY896321	98.60%
319	AY896319	97.24%
320	AY896316	97.59%
321	AY896313	96.93%
322	AY896309	96.93%
323	AY896307	97.60%

	324	AY896305		97.55%
	325	AY896303		96.93%
	326	AY896302		97.24%
	327	AY896301		96.63%
	328	AY896300		96.62%
	329	AY896299		97.24%
	330	AY896298		98.77%
	331	AY896297		97.22%
	332	AY896296		97.22%
	333	AY896295		97.22%
Cluster 12				
	1	EF568471	*	
Cluster 13				
	1	EF568468	*	
Cluster 14				
	1	EU916593	*	
	2	EU916529		100.00%
Cluster 15				
	1	AY896406	*	
	2	HQ660893		98.46%
	3	EF568512		98.46%
	4	EF568540		99.38%
	5	EF568486		99.07%
	6	EF568485		99.07%
	7	EF568479		99.07%
	8	EF568478		98.46%
	9	EF568475		99.07%
	10	HQ660942		98.46%
	11	HQ660928		99.07%
	12	HQ660926		98.77%
	13	HQ660909		99.07%
	14	HQ660906		99.07%
	15	HQ660903		98.77%
	16	HQ660888		99.07%
	17	HQ660873		98.50%
	18	HQ660869		99.07%
	19	HQ660858		98.77%
	20	HQ660853		99.07%
	21	HQ660812		99.07%
	22	HQ456009		98.77%
	23	HQ455985		99.38%
Cluster 16				
	1	HM210378	*	
	2	HM210384		97.55%
	3	HM210383		97.25%
	4	HM210380		97.55%
	5	HM210379		98.70%
	6	HM210377		98.47%
	7	HM210375		99.39%
	8	HM210374		97.25%
	9	HM210373		98.47%
	10	HM210372		96.64%
	11	HM210371		97.86%
	12	HM210370		97.25%
	13	HM210369		98.70%
	14	HM210365		96.64%
Cluster 17				
	1	KF151763	*	
Cluster 18				
	1	190192	*	
Cluster 19				
	1	EU916661	*	
	2	EU916655		99.08%
	3	EU916654		100.00%
	4	EU916653		99.69%
	5	EU916650		99.39%
	6	EU916648		100.00%
	7	EU916645		99.69%
	8	EU916641		99.69%
	9	EU916639		100.00%

	10	EU916637		99.69%
	11	EU916636		99.69%
	12	EU916635		99.69%
	13	EU916633		99.69%
	14	EU916631		99.39%
	15	EU916628		99.69%
	16	EU916627		99.69%
	17	EU916626		99.69%
	18	EU916625		99.08%
	19	EU916624		100.00%
	20	EU916622		99.69%
	21	EU916621		99.69%
	22	EU916620		100.00%
	23	EU916618		100.00%
	24	EU916615		100.00%
	25	EU916613		99.69%
	26	EU916611		100.00%
	27	EU916610		99.69%
	28	EU916609		99.69%
	29	EU916603		100.00%
	30	EU916597		99.69%
	31	EU916596		99.39%
	32	EU916591		99.69%
	33	EU916590		100.00%
	34	EU916586		100.00%
	35	EU916564		100.00%
	36	EU916556		99.69%
	37	EU916549		100.00%
	38	EU916546		100.00%
	39	EU916541		99.69%
	40	EU916539		100.00%
	41	EU916534		99.69%
	42	EU916533		100.00%
	43	EU916532		99.39%
	44	EU916531		100.00%
	45	EU916525		100.00%
Cluster 20				
	1	EU916326	*	
	2	EU916329		99.39%
Cluster 21				
	1	AY896371	*	
	2	HQ660814		98.70%
	3	DQ481336		98.70%
	4	DQ481335		99.08%
	5	DQ481334		99.08%
	6	DQ481333		99.39%
	7	DQ481331		99.08%
	8	DQ481330		98.78%
	9	DQ481328		99.08%
	10	DQ481327		99.08%
	11	DQ481325		98.78%
	12	DQ481323		98.78%
	13	KF619537		99.39%
	14	KF619536		99.08%
	15	KC013173		99.39%
	16	KC013059		99.08%
	17	KC013054		99.39%
	18	KC013052		98.78%
	19	KC013051		99.39%
	20	KC013050		99.39%
	21	KC013049		99.69%
	22	KC013047		99.39%
	23	EU052611		99.08%
	24	EU052555		99.08%
	25	EU052543		99.08%
	26	EU052529		99.39%
	27	EU052413		98.47%
	28	EU052412		99.39%
	29	EU052409		99.39%
	30	EU052408		99.08%

31	EU052406	98.47%
32	EU052405	98.78%
33	EU052380	98.78%
34	EU052350	98.47%
35	EU052343	99.08%
36	EU052342	99.08%
37	EU052329	99.08%
38	EU052327	98.78%
39	EU052319	99.08%
40	EU052297	98.70%
41	AY800143	97.55%
42	AY800142	98.70%
43	AY800140	96.94%
44	AY800139	98.78%
45	AY800138	99.69%
46	AY800137	99.08%
47	AY800136	98.78%
48	AY800135	98.47%
49	AY800134	99.39%
50	AB928259	99.08%
51	AB928254	98.78%
52	AB928253	98.78%
53	AB928242	99.39%
54	AB928236	99.08%
55	AB928220	99.08%
56	DQ825742	99.08%
57	DQ825739	99.38%
58	EF568515	99.39%
59	EF568521	99.08%
60	HQ611955	99.39%
61	HQ611954	99.69%
62	HQ611952	99.69%
63	HQ611940	99.69%
64	HQ611938	99.69%
65	HQ611926	99.39%
66	HQ611918	98.47%
67	HQ611912	99.08%
68	HQ611908	99.69%
69	HQ611893	99.69%
70	HQ611888	99.08%
71	HQ611887	98.78%
72	HQ611886	99.69%
73	HQ611884	99.08%
74	HQ611883	98.78%
75	HQ611882	99.08%
76	HQ611878	99.39%
77	HQ611873	99.69%
78	HQ611869	99.39%
79	HQ611867	99.39%
80	HQ611866	99.69%
81	HQ611857	99.08%
82	HQ611849	96.94%
83	HQ611848	99.69%
84	HQ611845	99.08%
85	HQ611840	99.39%
86	HQ611836	99.39%
87	HQ611834	99.69%
88	HQ611824	98.70%
89	HQ611817	98.70%
90	HQ611813	98.47%
91	HQ611810	97.55%
92	HQ611802	98.70%
93	HQ611798	98.47%
94	HQ611790	98.70%
95	HQ611788	99.69%
96	HQ611783	98.47%
97	HQ611781	98.47%
98	HQ611778	98.70%
99	HQ611777	98.78%
100	HQ611773	97.86%

101	HQ611769	98.70%
102	HQ611768	98.47%
103	HQ611761	98.47%
104	HQ611754	97.55%
105	HQ611750	98.47%
106	HQ611744	98.47%
107	HQ611740	98.47%
108	HQ611738	100.00%
109	HQ611737	98.70%
110	HQ611736	98.47%
111	HQ611733	98.47%
112	HQ611731	98.47%
113	HQ611730	97.86%
114	HQ611727	98.47%
115	HQ611724	98.47%
116	HQ611718	98.47%
117	HQ611717	97.86%
118	HQ611716	98.47%
119	HQ611713	98.70%
120	HQ611712	98.70%
121	HQ611711	98.70%
122	HQ611710	98.47%
123	HQ611708	98.70%
124	HQ611707	98.47%
125	HQ611701	98.47%
126	HQ611700	96.02%
127	HQ611698	100.00%
128	HQ611692	100.00%
129	HQ611691	99.69%
130	HQ611690	100.00%
131	HQ611686	99.08%
132	HQ611683	100.00%
133	HQ611681	100.00%
134	HQ611680	99.39%
135	HQ611679	99.69%
136	HQ611677	99.39%
137	HQ611676	99.08%
138	HQ611674	99.69%
139	HQ611672	99.69%
140	HQ611669	99.39%
141	HQ611668	99.39%
142	HQ611667	98.78%
143	HQ611665	99.69%
144	HQ611663	99.08%
145	HQ611658	99.69%
146	HQ611657	99.08%
147	HQ611651	99.69%
148	HQ611643	97.55%
149	HQ611641	98.70%
150	HQ611630	99.08%
151	HQ611627	99.69%
152	HQ611625	97.25%
153	HQ611619	99.08%
154	HQ611604	99.08%
155	HQ611603	98.70%
156	HQ611602	98.47%
157	HQ611600	98.78%
158	HQ611593	99.39%
159	HQ611591	99.08%
160	HQ611582	99.08%
161	HQ611578	99.08%
162	HQ611572	99.08%
163	HQ611554	99.69%
164	HQ611553	99.39%
165	HQ611551	99.39%
166	HQ611549	99.69%
167	HQ611547	99.39%
168	HQ611538	99.08%
169	HQ611537	99.39%
170	HQ611533	99.69%

171	HQ611515	99.39%
172	HQ611509	99.39%
173	HQ611504	98.78%
174	HQ611503	98.70%
175	HQ611500	99.08%
176	HQ611499	98.78%
177	HQ611498	98.78%
178	HQ611497	99.69%
179	HQ611495	99.69%
180	HQ611493	98.78%
181	HQ611490	99.69%
182	HQ611489	98.78%
183	HQ611488	99.69%
184	HQ611485	99.39%
185	HQ611482	99.39%
186	HQ611481	99.39%
187	HQ611472	99.39%
188	HQ611469	99.08%
189	HQ611468	99.39%
190	HQ611436	98.78%
191	HQ611431	98.70%
192	HQ611430	98.78%
193	HQ611426	97.55%
194	HQ611424	98.78%
195	HQ611415	98.47%
196	HQ611414	98.70%
197	HQ611413	99.39%
198	HQ611412	98.47%
199	HQ611409	98.78%
200	HQ611408	98.47%
201	HQ611406	98.47%
202	HQ611405	98.70%
203	HQ611404	99.08%
204	HQ611401	98.78%
205	HQ611400	98.78%
206	HQ611398	98.70%
207	HQ611396	98.47%
208	HQ611395	98.70%
209	HQ611393	99.39%
210	HQ611392	98.47%
211	HQ611390	98.70%
212	HQ611385	98.70%
213	HQ611380	99.08%
214	HQ611378	98.70%
215	HQ611377	98.78%
216	HQ611376	99.08%
217	HQ611371	99.69%
218	HQ611369	98.78%
219	HQ611368	98.78%
220	HQ611367	98.47%
221	HQ611365	99.39%
222	HQ611364	99.69%
223	HQ611362	98.78%
224	HQ611361	99.69%
225	HQ611360	99.08%
226	HQ611359	98.78%
227	HQ611357	99.69%
228	HQ611356	98.78%
229	HQ611354	99.69%
230	HQ611353	98.47%
231	AB727486	99.39%
232	AB727482	99.08%
233	AB727481	99.69%
234	AB727478	99.39%
235	AB727472	99.69%
236	AB727471	99.39%
237	AB727470	99.39%
238	HQ456121	99.39%
239	HQ456116	99.39%
240	HQ456115	98.78%

	241	HQ456110		99.08%
	242	HQ456108		99.69%
	243	HQ456105		99.69%
	244	HQ456099		98.78%
	245	HQ456098		99.08%
	246	HQ456096		100.00%
	247	HQ456095		99.08%
	248	HQ456045		100.00%
	249	HQ456041		100.00%
	250	HQ456019		99.39%
	251	HQ456013		98.78%
	252	HQ456012		98.78%
	253	HQ456001		99.08%
	254	HQ455999		99.39%
	255	HQ455998		99.08%
	256	HQ455996		99.08%
	257	HQ455991		99.08%
	258	HQ455990		98.78%
	259	HQ455989		99.39%
	260	HQ455979		98.78%
	261	HQ455978		100.00%
	262	HQ455968		99.39%
	263	HQ455967		99.39%
	264	HQ455966		99.39%
	265	HQ455964		99.69%
	266	HQ455962		98.47%
	267	HQ455961		98.78%
	268	HQ455960		99.08%
	269	HQ455930		99.39%
	270	HQ455918		99.69%
	271	HQ455917		99.39%
	272	HQ455909		98.47%
	273	HQ455908		100.00%
	274	HQ455902		99.39%
	275	HQ455891		98.78%
	276	HQ455890		99.39%
	277	HQ455883		99.69%
	278	GQ475478		98.78%
	279	AY896372		99.39%
	280	AY896370		99.70%
	281	AY896369		99.70%
	282	AY896334		99.09%
	283	AY896330		98.78%
	284	AY896329		99.39%
	285	AY896320		96.86%
	286	AY896318		99.39%
	287	AY896315		99.39%
	288	AY896314		97.87%
	289	AY896312		96.59%
	290	AY896311		99.69%
	291	AY896308		98.80%
	292	AY896306		99.38%
	293	AY896304		98.44%
	294	EF204556		98.70%
Cluster 22				
	1	EU916456	*	
	2	EU916454		99.39%
	3	EU916450		100.00%
	4	EU916440		99.08%
	5	EU916431		99.69%
	6	EU916423		99.69%
	7	EU916316		99.69%
	8	EU916306		99.69%
	9	EU916301		99.69%
	10	EU916287		99.69%
	11	EU916286		100.00%
	12	EU916280		99.08%
Cluster 23				
	1	KF151389		97.25%
	2	HQ611745		97.25%

	3	EU916327	*	
	4	EU916323		96.33%
	5	KC140395		97.25%
	6	KC140394		96.94%
	7	KC140393		97.25%
	8	KC140392		97.25%
	9	KC140391		97.55%
	10	KC140367		96.94%
	11	KC140366		97.55%
	12	KC140363		97.25%
	13	KC140362		97.25%
	14	KC140360		97.25%
	15	KC140359		97.55%
	16	KC140358		97.25%
	17	KC140357		97.55%
	18	KC140355		97.55%
	19	KC140395		97.25%
	20	KC140394		96.94%
	21	KC140393		97.25%
	22	KC140392		97.25%
	23	KC140391		97.55%
	24	KC140367		96.94%
	25	KC140366		97.55%
	26	KC140363		97.25%
	27	KC140362		97.25%
	28	KC140360		97.25%
	29	KC140359		97.55%
	30	KC140358		97.25%
	31	KC140357		97.55%
	32	KC140355		97.55%
Cluster 24				
	1	GQ475477	*	
Cluster 25				
	1	AY896462	*	
	2	AY896455		99.70%
Cluster 26				
	1	AY896456	*	
	2	DQ481326		99.69%
	3	HQ456044		96.02%
	4	HQ456033		96.02%
	5	HQ456031		96.02%
Cluster 27				
	1	EF204560	*	
Cluster 28				
	1	HM801610	*	
	2	HM801752		99.69%
	3	HM801741		99.39%
	4	HM801737		99.39%
Cluster 29				
	1	HM801457	*	
	2	DQ481274		97.25%
	3	DQ481273		98.47%
	4	DQ481270		99.39%
	5	HM210414		98.78%
	6	HM210411		96.94%
	7	HM210407		97.55%
	8	HM210404		98.70%
	9	HM210364		98.47%
	10	HM210361		98.47%
	11	HM210358		99.08%
	12	HM210342		99.08%
	13	HM210311		96.94%
	14	HM801572		98.78%
	15	HM801570		98.47%
	16	HM801568		98.47%
	17	HM801562		98.47%
	18	HM801553		98.47%
	19	HM801550		98.47%
	20	HM801548		98.47%
	21	HM801547		98.47%

	22	HM801545	98.47%
	23	HM801543	98.47%
	24	HM801541	98.47%
	25	HM801537	98.47%
	26	HM801536	98.47%
	27	HM801534	98.78%
	28	HM801533	98.47%
	29	HM801532	98.78%
	30	HM801527	98.70%
	31	HM801524	97.86%
	32	HM801523	98.47%
	33	HM801521	98.78%
	34	HM801520	98.47%
	35	HM801516	98.47%
	36	HM801460	98.78%
	37	HM801452	98.78%
	38	HM801451	98.78%
	39	HM801445	98.78%
	40	HM801443	99.08%
	41	HM801438	98.78%
	42	HM801437	98.47%
	43	HM801426	98.78%
	44	HM801416	98.70%
	45	HM801415	98.47%
	46	HM801397	98.78%
	47	HM801382	98.47%
	48	HM801381	98.47%
	49	HM801364	97.86%
	50	HM801360	98.47%
	51	HM801356	99.08%
	52	HM801354	98.78%
	53	HM801346	98.70%
	54	HM801344	98.78%
	55	KF151665	98.78%
	56	KF151480	98.78%
Cluster 30			
	1	HM801743	*
	2	HM801751	99.08%
	3	HQ611583	100.00%
Cluster 31			
	0	HM801623	*
	1	HM801708	98.47%
	2	6666666.713	98.78%
Cluster 32			
	1	HM801245	*
	2	HM801277	99.08%
	3	HM801272	99.39%
	4	HM801265	99.08%
	5	HM801264	99.08%
	6	HM801261	98.78%
	7	HM801259	99.08%
	8	HM801257	99.08%
	9	HM801250	98.47%
Cluster 33			
	1	HM801744	*
	2	HM801771	96.02%
	3	HM801770	96.02%
	4	HM801769	96.02%
	5	HM801767	96.02%
Cluster 34			
	1	HM801228	*
	2	HM801239	99.08%
	3	HM801233	98.78%
	4	HM801225	99.39%
Cluster 35			
	1	KF151844	*
	2	KF151843	99.39%
	3	KF151840	99.69%
	4	KF151838	99.69%
	5	KF151837	99.69%

	6	KF151806		96.94%
	7	KF151800		96.64%
	8	KF151790		96.64%
	9	KF151784		99.08%
	10	KF151779		96.64%
	11	KF151776		99.69%
	12	KF151775		99.39%
	13	KF151772		97.25%
	14	KF151771		99.69%
	15	KF151769		99.69%
	16	KF151768		99.69%
	17	KF151767		99.39%
	18	KF151766		99.39%
	19	KF151670		99.69%
	20	KF151621		96.94%
	21	KF151613		99.69%
	22	KF151606		99.69%
	23	KF151599		99.39%
	24	KF151592		99.08%
Cluster 36				
	1	KF151711	*	
	2	KF151719		99.08%
Cluster 37				
	1	KF151482	*	
Cluster 38				
	1	KF151590	*	
	2	KF151591		99.39%
	3	KF151585		99.08%
	4	KF151582		98.78%
Cluster 39				
	1	KF151699	*	
	2	KF151706		99.08%
	3	KF151703		99.39%
	4	KF151702		99.69%
	5	KF151701		99.39%
	6	KF151700		99.39%
	7	KF151696		99.39%
	8	KF151695		99.39%
	9	KF151694		99.39%
Cluster 40				
	1	KF151523	*	
	2	KF151524		99.69%
	3	KF151522		99.39%
	4	KF151520		99.69%
	5	KF151519		99.69%
	6	KF151517		99.39%
	7	KF151516		99.69%
Cluster 41				
	1	KF151610	*	
	2	KF151609		99.69%
Cluster 42				
	1	KF151587	*	
Cluster 43				
	1	KF151583	*	
Cluster 44				
	1	HQ660894	*	
	2	HQ660938		99.08%
	3	HQ660925		99.39%
	4	HQ660922		98.78%
	5	HQ660889		98.78%
	6	HQ660877		98.78%
	7	HQ660857		99.08%
Cluster 45				
	1	HQ660860	*	
Cluster 46				
	1	HQ660890	*	
Cluster 47				
	1	HQ660866	*	
Cluster 48				
	1	HQ660897	*	

	2	HQ660935		96.33%
	3	HQ660934		96.02%
	4	HQ660919		96.02%
	5	HQ660898		96.94%
Cluster 49				
	1	HQ660899	*	
Cluster 50				
	1	HQ660936	*	
Cluster 51				
	1	HQ660876	*	
Cluster 52				
	1	HQ660883	*	
Cluster 53				
	1	HQ660918	*	
	2	HQ660854		99.69%
Cluster 54				
	1	HQ660904	*	
	2	HQ660915		96.33%
Cluster 55				
	1	HQ660916	*	
	2	HQ660880		96.02%
Cluster 56				
	1	HQ660917	*	
Cluster 57				
	1	HQ660875	*	
Cluster 58				
	1	HQ660905	*	
	2	HQ660914		99.69%
Cluster 59				
	1	HQ660927	*	
	2	HQ660940		96.33%
Cluster 60				
	1	HQ660881	*	
Cluster 61				
	1	HQ660908	*	
	2	HQ660861		98.78%
	3	HQ660859		99.39%
Cluster 62				
	1	HQ660855	*	
Cluster 63				
	1	HQ660870	*	
Cluster 64				
	1	HQ660879	*	
Cluster 65				
	1	HQ660941	*	
Cluster 66				
	1	HQ660885	*	
Cluster 67				
	1	HQ660920	*	
	2	HQ660937		99.69%
Cluster 68				
	1	HQ660902	*	
	2	HQ660907		98.78%
Cluster 69				
	1	HQ660887	*	
	2	HQ660892		99.69%
Cluster 70				
	1	HQ660895	*	
Cluster 71				
	1	HQ660867	*	
Cluster 72				
	1	HQ660872	*	
Cluster 73				
	1	JQ358703	*	
	2	JQ358700		98.70%
	3	JQ358699		99.69%
	4	JQ358695		99.69%
	5	JQ358692		99.69%
	6	JQ358691		99.08%
	7	JQ358675		99.69%

	8	JQ358672	99.69%
	9	JQ358643	99.69%
	10	HQ611804	98.47%
	11	HQ611784	98.78%
	12	HQ611654	98.78%
	13	HQ611645	98.78%
	14	HQ611458	99.39%
	15	HQ611455	99.39%
	16	HQ611452	98.78%
	17	HQ611449	99.39%
Cluster 74			
	1	JQ358656	*
	2	HM210334	98.78%
	3	KF151720	99.69%
	4	KF151718	99.39%
	5	KF151716	99.69%
	6	KF151713	99.39%
	7	KF151709	99.69%
	8	KF151707	99.69%
	9	KF151704	99.69%
	10	KF151680	99.39%
	11	KF151676	99.69%
	12	KF151663	99.69%
	13	KF151662	99.69%
	14	KF151658	99.08%
	15	KF151655	99.69%
	16	KF151654	99.39%
	17	KF151651	99.69%
	18	KF151649	99.69%
	19	KF151648	99.69%
	20	KF151647	99.69%
	21	KF151645	99.69%
	22	KF151644	99.69%
	23	KF151643	99.39%
	24	KF151635	99.69%
	25	KF151634	99.08%
	26	KF151633	99.69%
	27	KF151632	99.39%
	28	KF151629	99.08%
	29	KF151628	99.69%
	30	KF151625	99.69%
	31	KF151579	99.69%
	32	KF151577	99.69%
	33	KF151576	99.39%
	34	KF151575	99.39%
	35	KF151573	99.69%
	36	KF151572	99.69%
	37	KF151571	99.69%
	38	KF151570	99.39%
	39	KF151569	99.69%
	40	KF151566	99.69%
	41	KF151565	99.39%
	42	KF151564	99.69%
	43	KF151562	99.69%
	44	KF151558	99.39%
	45	KF151557	99.69%
	46	KF151556	99.08%
	47	KF151463	99.69%
	48	KF151449	99.69%
	49	HQ611953	99.69%
	50	HQ611950	100.00%
	51	HQ611948	99.69%
	52	HQ611946	99.69%
	53	HQ611944	100.00%
	54	HQ611943	99.69%
	55	HQ611942	99.69%
	56	HQ611941	100.00%
	57	HQ611923	100.00%
	58	HQ611914	99.08%
	59	HQ611900	99.39%

60	HQ611885	99.69%
61	HQ611872	99.39%
62	HQ611870	100.00%
63	HQ611863	99.69%
64	HQ611862	99.69%
65	HQ611860	100.00%
66	HQ611841	100.00%
67	HQ611829	99.69%
68	HQ611828	100.00%
69	HQ611825	100.00%
70	HQ611821	99.69%
71	HQ611820	100.00%
72	HQ611816	99.69%
73	HQ611812	100.00%
74	HQ611801	99.69%
75	HQ611800	99.69%
76	HQ611794	100.00%
77	HQ611792	100.00%
78	HQ611766	100.00%
79	HQ611760	99.08%
80	HQ611757	100.00%
81	HQ611756	99.69%
82	HQ611732	99.69%
83	HQ611699	100.00%
84	HQ611696	99.69%
85	HQ611695	99.69%
86	HQ611694	100.00%
87	HQ611693	99.39%
88	HQ611689	100.00%
89	HQ611688	99.69%
90	HQ611685	99.39%
91	HQ611684	100.00%
92	HQ611678	99.69%
93	HQ611675	99.08%
94	HQ611673	99.69%
95	HQ611670	100.00%
96	HQ611666	100.00%
97	HQ611664	100.00%
98	HQ611661	100.00%
99	HQ611656	100.00%
100	HQ611655	99.69%
101	HQ611649	99.08%
102	HQ611640	100.00%
103	HQ611638	99.69%
104	HQ611636	99.69%
105	HQ611634	100.00%
106	HQ611632	100.00%
107	HQ611629	99.69%
108	HQ611628	99.39%
109	HQ611621	100.00%
110	HQ611618	100.00%
111	HQ611616	100.00%
112	HQ611610	99.69%
113	HQ611601	99.69%
114	HQ611592	100.00%
115	HQ611588	99.08%
116	HQ611585	99.69%
117	HQ611584	100.00%
118	HQ611576	99.69%
119	HQ611575	99.69%
120	HQ611565	99.69%
121	HQ611556	99.69%
122	HQ611552	100.00%
123	HQ611548	99.69%
124	HQ611546	100.00%
125	HQ611543	99.39%
126	HQ611539	99.69%
127	HQ611530	99.69%
128	HQ611528	100.00%
129	HQ611512	100.00%

	130	HQ611510		99.39%
	131	HQ611506		100.00%
	132	HQ611502		99.69%
	133	HQ611496		99.69%
	134	HQ611492		99.69%
	135	HQ611457		99.39%
	136	HQ611456		100.00%
	137	HQ611454		99.69%
	138	HQ611440		99.69%
	139	HQ611437		99.39%
Cluster 75				
	1	JQ358650	*	
Cluster 76				
	1	JQ358634	*	
	2	JQ358633		99.39%
Cluster 77				
	1	JF429970		100.00%
	2	JF429969	*	
Cluster 78				
	1	JF429967	*	
Cluster 79				
	1	KF151412	*	
Cluster 80				
	1	KF151411	*	
Cluster 81				
	1	KF151410	*	
Cluster 82				
	1	KF151409	*	
Cluster 83				
	1	KF151408	*	
Cluster 84				
	1	KF151407	*	
Cluster 85				
	1	KF151406	*	
Cluster 86				
	1	KF151405	*	
Cluster 87				
	1	KF151404	*	
Cluster 88				
	1	KF151403	*	
Cluster 89				
	1	KF151402	*	
Cluster 90				
	1	KF151401	*	
Cluster 91				
	1	KF151400	*	
Cluster 92				
	1	KF151399	*	
Cluster 93				
	1	KF151398	*	
Cluster 94				
	1	KF151397	*	
	2	KF151394		96.64%
Cluster 95				
	1	KF151396	*	
Cluster 96				
	1	KF151395	*	
Cluster 97				
	1	KF151393	*	
Cluster 98				
	1	KF151392	*	
Cluster 99				
	1	KF151391	*	
Cluster 100				
	1	KF151390	*	
Cluster 101				
	1	AY191971	*	
	2	JF429955		99.38%
	3	EU052658		99.07%
	4	EU052655		98.46%

5	EU052650	99.07%
6	EU052649	99.07%
7	EU052648	99.07%
8	EU052647	99.69%
9	EU052645	99.07%
10	EU052643	98.46%
11	EU052642	98.77%
12	EU052641	99.07%
13	EU052636	99.07%
14	EU052633	99.07%
15	EU052632	98.77%
16	EU052631	99.07%
17	EU052630	99.07%
18	EU052628	99.07%
19	EU052622	98.46%
20	EU052621	99.07%
21	EU052620	98.77%
22	EU052618	99.07%
23	EU052617	99.07%
24	EU052616	98.77%
25	EU052615	99.07%
26	EU052614	98.77%
27	EU052613	99.07%
28	EU052612	99.07%
29	EU052610	99.07%
30	EU052608	98.77%
31	EU052607	98.77%
32	EU052606	98.77%
33	EU052605	98.46%
34	EU052603	99.07%
35	EU052602	98.77%
36	EU052600	98.46%
37	EU052599	98.46%
38	EU052597	98.50%
39	EU052576	99.07%
40	EU052517	98.77%
41	EU052516	99.07%
42	EU052512	99.07%
43	EU052508	98.77%
44	EU052506	99.07%
45	EU052505	99.07%
46	EU052504	99.07%
47	EU052503	98.77%
48	EU052500	98.77%
49	EU052499	98.46%
50	EU052491	99.07%
51	EU052488	99.07%
52	EU052485	99.07%
53	EU052480	99.07%
54	EU052479	99.07%
55	EU052478	99.07%
56	EU052477	99.07%
57	EU052476	99.07%
58	EU052473	99.07%
59	EU052472	99.07%
60	EU052471	99.07%
61	EU052470	97.84%
62	EU052466	98.46%
63	EU052465	99.07%
64	EU052463	99.07%
65	EU052461	99.07%
66	EU052460	98.77%
67	EU052456	99.07%
68	EU052455	99.07%
69	EU052453	99.07%
70	EU052452	99.07%
71	EU052450	98.46%
72	EU052447	99.07%
73	EU052445	98.77%
74	EU052442	99.38%

	75	EU052441	99.07%
	76	EU052439	99.07%
	77	EU052392	99.07%
	78	EU052391	98.50%
	79	EU052389	89.50%
	80	EU052387	98.46%
	81	EU052386	99.38%
	82	EU052385	99.07%
	83	EU052384	99.07%
	84	EU052383	99.07%
	85	EU052382	99.07%
	86	DQ825748	99.07%
	87	HQ455933	99.38%
	88	HQ455931	99.38%
Cluster 102			
	1	AY191961	*
	2	DQ481271	96.02%
	3	DQ481265	96.02%
	4	DQ481264	96.02%
Cluster 103			
	1	DQ481390	*
	2	DQ481365	99.69%
	3	DQ825731	99.39%
	4	EF204558	96.02%
Cluster 104			
	1	DQ481329	*
	2	HQ456047	98.47%
	3	HQ456043	987.00%
Cluster 105			
	1	DQ481322	*
	2	DQ481288	987.00%
	3	DQ481285	96.94%
	4	DQ481283	97.86%
	5	DQ481281	97.86%
	6	HQ455940	98.78%
Cluster 106			
	1	DQ481320	*
	2	DQ481314	99.08%
	3	DQ481289	98.47%
	4	DQ481287	99.08%
	5	DQ481286	99.08%
	6	DQ481284	99.08%
	7	DQ481282	98.47%
	8	DQ481280	98.50%
	9	DQ481279	99.08%
	10	DQ481278	98.47%
Cluster 107			
	1	DQ481319	*
	2	DQ481313	99.39%
Cluster 108			
	1	DQ481277	*
	2	DQ481267	96.02%
	3	DQ481266	96.33%
Cluster 109			
	1	DQ481276	*
	2	DQ481275	96.64%
	3	DQ481269	96.02%
	4	DQ481268	96.02%
	5	DQ481263	96.02%
Cluster 110			
	1	DQ481272	*
Cluster 111			
	1	HQ229024	*
Cluster 112			
	1	HQ229023	*
Cluster 113			
	1	HQ229019	*
Cluster 114			
	1	HM210413	*
	2	HM210412	97.55%

	3	HM210360		96.33%
Cluster 115				
	1	HM210409	*	
	2	GQ475450		98.78%
	3	GQ475448		99.39%
	4	AY896431		98.70%
Cluster 116				
	1	HM210408	*	
	2	HM210376		97.55%
	3	HM210359		987.00%
Cluster 117				
	1	HM210406	*	
Cluster 118				
	1	HM210405	*	
Cluster 119				
	1	HM210403	*	
	2	HM210396		99.08%
	3	HM210394		99.69%
	4	HM210393		99.08%
	5	HM210388		98.47%
	6	HM210386		99.08%
	7	KC013226		99.39%
	8	KC013223		98.78%
	9	KC013175		99.08%
	10	KC013172		99.08%
	11	HQ611910		97.55%
	12	HQ611909		97.86%
	13	HQ611880		96.94%
	14	HQ611839		96.64%
	15	HQ611833		96.02%
	16	HQ611831		96.64%
	17	HQ611808		96.94%
	18	HQ611797		96.94%
	19	HQ611785		96.33%
	20	HQ611780		96.33%
	21	HQ611774		96.94%
	22	HQ611765		97.86%
	23	HQ611764		96.94%
	24	HQ611762		96.02%
	25	HQ611755		96.94%
	26	HQ611748		96.94%
	27	HQ611747		96.64%
	28	HQ611739		96.94%
	29	HQ611646		96.94%
	30	HQ611644		96.64%
	31	HQ611637		96.94%
	32	HQ611590		96.33%
	33	HQ611564		98.78%
	34	HQ611480		98.70%
	35	HQ611474		99.08%
	36	HQ611473		98.70%
	37	HQ456117		99.08%
	38	HQ456107		99.08%
	39	HQ456106		99.08%
	40	HQ456102		99.39%
Cluster 120				
	1	HM210395	*	
	2	KF151762		98.78%
Cluster 121				
	1	HM210390	*	
Cluster 122				
	1	HM210385	*	
	2	HM210381		97.86%
	3	HM210366		97.25%
Cluster 123				
	1	HM210367	*	
Cluster 124				
	1	HM210362	*	
Cluster 125				
	1	HM210345	*	

Cluster 126			
	1	HM210344	*
	2	HM210317	99.08%
	3	HQ455868	96.94%
	4	HQ455864	98.47%
	5	EU916317	96.94%
	6	EU916311	96.94%
Cluster 127			
	1	HM210343	*
	2	HM210341	97.25%
	3	HM210339	97.86%
	4	HM210338	98.78%
	5	HM210335	96.33%
	6	HM210333	98.70%
	7	HM210332	98.47%
	8	HM210331	96.94%
	9	HM210330	96.64%
	10	HM210329	96.33%
	11	HM210328	98.70%
	12	HM210326	96.94%
	13	HM210325	98.78%
	14	HM210324	99.08%
	15	HM210316	96.02%
Cluster 128			
	1	HM210340	*
Cluster 129			
	1	HM210336	*
Cluster 130			
	1	HM210327	*
Cluster 131			
	1	HM210323	*
Cluster 132			
	1	HM210320	*
Cluster 133			
	1	HM210319	*
Cluster 134			
	1	HM210318	*
Cluster 135			
	1	HM210315	*
Cluster 136			
	1	HM210314	*
	2	HM210313	96.64%
	3	HM210312	96.94%
	4	HM210310	97.55%
Cluster 137			
	1	HM210309	*
	2	HM801583	96.02%
	3	HM801582	96.02%
	4	HM801581	96.02%
	5	HM801580	96.02%
	6	HM801578	96.02%
	7	HM801577	96.02%
	8	HM801576	96.02%
	9	HM801575	96.02%
	10	HM801574	96.33%
	11	HM801573	96.02%
	12	HM801569	96.02%
	13	HM801564	96.02%
	14	HM801563	96.02%
	15	HM801551	96.02%
	16	HM801539	96.02%
	17	HM801526	96.02%
	18	HM801519	96.02%
	19	HM801518	96.02%
	20	HM801515	96.02%
	21	HM801464	96.02%
	22	HM801462	96.02%
	23	HM801455	96.02%
	24	HM801454	96.02%
	25	HM801444	96.02%

	26	HM801440		96.02%
	27	HM801439		96.02%
	28	HM801434		96.02%
	29	HM801430		96.02%
	30	HM801424		96.02%
	31	HM801421		96.02%
	32	HM801420		96.02%
	33	HM801417		96.02%
	34	HM801411		96.02%
	35	HM801351		96.02%
	36	HM801349		96.02%
	37	HM801336		96.02%
	38	HM801226		96.33%
	39	KF151660		96.02%
Cluster 138				
	1	EU052660	*	
	2	EU052659		99.39%
	3	EU052656		98.78%
	4	EU052652		99.69%
Cluster 139				
	1	EU052627	*	
	2	HQ660911		96.64%
	3	HQ660901		96.94%
Cluster 140				
	1	EU052625	*	
Cluster 141				
	1	EU052594	*	
	2	EU052530		96.33%
	3	EU052524		98.78%
	4	HQ456075		98.78%
	5	HQ455867		98.47%
	6	HQ455852		98.47%
Cluster 142				
	1	EU052592	*	
	2	EU052526		99.69%
Cluster 143				
	1	EU052591	*	
Cluster 144				
	1	EU052584	*	
	2	EU052522		98.78%
	3	EU052321		99.39%
	4	GQ475443		99.08%
Cluster 145				
	1	EU052578	*	
	2	EU052538		96.64%
	3	EU052438		98.78%
	4	HQ456088		97.86%
Cluster 146				
	1	EU052577	*	
Cluster 147				
	1	EU052568	*	
Cluster 148				
	1	EU052562	*	
Cluster 149				
	1	EU052550	*	
Cluster 150				
	1	EU052549	*	
Cluster 151				
	1	EU052537	*	
Cluster 152				
	1	EU052532	*	
Cluster 153				
	1	EU052531	*	
	2	EU052323		98.78%
Cluster 154				
	1	EU052523	*	
Cluster 155				
	1	EU052403	*	
Cluster 156				
	1	EU052395	*	

Cluster 157				
	1	EU052335	*	
Cluster 158				
	1	EU052326	*	
Cluster 159				
	1	EU052325	*	
Cluster 160				
	1	EU052324	*	
	2	GQ475428		96.64%
Cluster 161				
	1	EU052322	*	
Cluster 162				
	1	AY800141	*	
Cluster 163				
	1	AB928289	*	
	2	KF151836		98.47%
	3	KF151835		98.78%
	4	KF151834		98.47%
	5	KF151831		98.47%
	6	KF151828		98.47%
	7	KF151826		98.70%
	8	KF151825		98.47%
	9	KF151824		98.70%
	10	KF151823		98.47%
	11	KF151822		98.47%
	12	KF151821		98.47%
	13	KF151819		98.70%
	14	KF151816		98.47%
	15	KF151815		97.86%
	16	KF151814		97.86%
	17	KF151813		98.47%
	18	KF151812		98.47%
	19	KF151811		98.70%
	20	KF151751		98.47%
	21	KF151749		98.47%
	22	KF151748		98.47%
	23	KF151742		98.47%
	24	KF151740		98.70%
	25	KF151736		98.47%
	26	KF151734		98.70%
	27	KF151513		98.47%
	28	KF151512		97.86%
	29	KF151510		98.70%
	30	KF151507		98.47%
	31	KF151504		98.70%
	32	KF151500		97.86%
	33	KF151491		97.86%
	34	KF151490		98.47%
	35	KF151489		98.47%
	36	KF151487		98.47%
	37	KF151484		98.47%
	38	HQ611949		98.78%
	39	HQ611864		98.78%
	40	HQ611851		98.78%
	41	EF204562		99.69%
Cluster 164				
	1	AB928233	*	
	2	AB928232		99.69%
	3	Vibrio diazotrophicus		96.33%
	4	1348635.5		97.25%
	5	HQ456025		96.94%
	6	HQ455901		98.47%
	7	HQ455881		98.47%
	8	HQ455875		98.47%
	9	HQ455843		98.47%
Cluster 165				
	1	EU151794	*	
	2	EU151793		98.70%
	3	EU151792		98.78%
	4	EU151791		98.47%

	5	EU151790		99.39%
	6	EU151789		98.78%
Cluster 166				
	1	EU151788	*	
	2	1122201.7		99.08%
	3	HQ456042		97.55%
	4	HQ455836		97.86%
Cluster 167				
	1	EU151787	*	
Cluster 168				
	1	EU151785	*	
Cluster 169				
	1	EU151784	*	
Cluster 170				
	1	EU151782	*	
	2	EU151775		99.69%
	3	HQ456026		96.64%
	4	HQ456017		96.94%
	5	HQ456016		96.64%
	6	HQ456007		96.33%
	7	HQ456006		96.94%
	8	HQ455942		96.33%
	9	HQ455913		96.94%
	10	HQ455870		96.94%
	11	HQ455857		97.25%
Cluster 171				
	1	EU151781	*	
Cluster 172				
	1	EU151780	*	
Cluster 173				
	1	EU151779	*	
Cluster 174				
	1	EU151778	*	
Cluster 175				
	1	EU151777	*	
Cluster 176				
	1	EU151776	*	
Cluster 177				
	1	EU151774	*	
Cluster 178				
	1	EU151773	*	
Cluster 179				
	1	DQ825752	*	
Cluster 180				
	1	DQ825750	*	
Cluster 181				
	1	DQ825743	*	
Cluster 182				
	1	DQ825737	*	
	2	HQ455869		96.64%
Cluster 183				
	1	DQ825735	*	
Cluster 184				
	1	DQ825726	*	
Cluster 185				
	1	DQ825727		99.38%
	2	DQ825725	*	
Cluster 186				
	1	DQ825724	*	
Cluster 187				
	1	DQ825722	*	
Cluster 188				
	1	DQ825721	*	
Cluster 189				
	1	DQ825718	*	
	2	680279.4		97.86%
Cluster 190				
	1	DQ825714	*	
Cluster 191				
	1	DQ825713	*	

	2	DQ825712		99.08%
Cluster 192				
	1	EF568590	*	
	2	EF568589		99.39%
	3	EF568587		99.08%
	4	EF568580		99.08%
	5	EF568574		99.08%
Cluster 193				
	1	EF568563	*	
Cluster 194				
	1	EF568559	*	
Cluster 195				
	1	EF568558	*	
	2	HQ455853		97.55%
Cluster 196				
	1	EF568555	*	
Cluster 197				
	1	EF568551	*	
Cluster 198				
	1	EF568550	*	
Cluster 199				
	1	EF568549	*	
	2	EF568546		99.08%
Cluster 200				
	1	EF568548	*	
Cluster 201				
	1	EF568545	*	
Cluster 202				
	1	EF568544	*	
Cluster 203				
	1	EF568536	*	
	2	EF568534		99.69%
	3	EF568531		99.69%
Cluster 204				
	1	EF568532	*	
Cluster 205				
	1	EF568530	*	
Cluster 206				
	1	EF568474	*	
	2	EF568467		99.69%
Cluster 207				
	1	EF568441	*	
Cluster 208				
	1	EF568439	*	
	2	EF568566		99.69%
Cluster 209				
	1	EF568428	*	
	2	EF568427		99.08%
	3	EF568425		99.39%
Cluster 210				
	1	EF568426	*	
	2	EF568423		99.69%
	3	EF568420		99.39%
	4	EF568419		99.69%
	5	EF568416		99.39%
	6	EF568415		99.69%
	7	EF568414		99.69%
Cluster 211				
	1	EF568422	*	
	2	EF568417		99.69%
Cluster 212				
	1	EF568605	*	
	2	EF568603		99.69%
	3	EF568599		99.69%
	4	EF568593		99.69%
	5	EF568591		99.69%
Cluster 213				
	1	EF568437	*	
Cluster 214				
	1	EF568435	*	

Cluster 215			
	1	EF568434	*
	2	EF568429	99.39%
Cluster 216			
	1	HM042891	*
Cluster 217			
	1	HM042890	*
	2	HM042889	99.39%
	3	HM042888	99.39%
	4	HM042885	99.39%
	5	HM042883	99.39%
	6	HM042879	99.08%
	7	Klebsiella oxytoca	96.33%
Cluster 218			
	1	HM042886	*
Cluster 219			
	1	Desulfovibrio_vulga	*
	2	Desulfovibrio	100.00%
Cluster 220			
	1	2Desulfobacter_latu	*
Cluster 221			
	1	3Desulfobacter_curv	*
Cluster 222			
	1	Chlorobium_limicola	*
	2	290315.4	99.69%
Cluster 223			
	1	Pelodictyon_luteolum	*
	2	Pelodictyon	99.69%
Cluster 224			
	1	HM801765	*
	2	HM801764	99.08%
	3	HM801763	99.39%
	4	HM801762	99.08%
	5	HM801761	98.78%
	6	HM801759	99.08%
	7	HM801758	99.08%
	8	HM801757	99.08%
	9	HM801756	99.08%
	10	HM801755	98.78%
	11	HM801754	99.08%
	12	HM801749	98.78%
	13	HM801748	99.08%
	14	HM801747	99.69%
	15	HM801745	99.08%
	16	HM801740	99.39%
	17	HM801739	99.08%
	18	HM801738	99.08%
Cluster 225			
	1	KF151830	*
	2	KF151829	99.69%
	3	KF151732	99.39%
	4	KF151731	99.69%
	5	KF151730	99.69%
	6	KF151727	99.39%
	7	KF151726	99.69%
	8	KF151725	99.69%
	9	KF151724	99.69%
	10	KF151511	99.69%
	11	KF151509	99.69%
	12	KF151503	99.69%
	13	KF151498	99.69%
	14	KF151486	99.39%
Cluster 226			
	1	KF151805	*
	2	KF151804	99.39%
	3	KF151799	99.08%
	4	KF151792	99.08%
	5	KF151791	98.47%
Cluster 227			
	1	KF151756	*

	2	KF151479		97.55%
Cluster 228				
	1	KF151721	*	
Cluster 229				
	1	KF151674	*	
	2	KF151673		99.69%
	3	KF151672		99.39%
Cluster 230				
	1	KF151656	*	
Cluster 231				
	1	KF151646	*	
	2	KF151551		99.69%
	3	KF151546		99.69%
	4	KF151545		99.69%
	5	KF151544		99.39%
	6	KF151540		99.69%
	7	KF151539		99.69%
	8	KF151537		99.69%
	9	KF151533		99.69%
	10	KF151528		99.69%
	11	HQ456114		987.00%
	12	HQ456112		97.86%
	13	HQ456111		97.25%
	14	HQ456109		987.00%
	15	HQ456104		97.86%
	16	HQ456101		97.86%
Cluster 232				
	1	KF151549	*	
Cluster 233				
	1	KF151548	*	
	2	KF151531		99.39%
Cluster 234				
	1	KF151527	*	
Cluster 235				
	1	KF151481	*	
Cluster 236				
	1	HQ660930	*	
Cluster 237				
	1	HQ660874		100.00%
	2	HQ660874	*	
Cluster 238				
	1	HQ660868	*	
Cluster 239				
	1	HQ660865	*	
Cluster 240				
	1	1121411.5	*	
	2	HQ611626		97.25%
Cluster 241				
	1	1121447.5	*	
Cluster 242				
	1	1121918.6	*	
Cluster 243				
	1	1235834.8	*	
Cluster 244				
	1	1294021.6	*	
Cluster 245				
	1	1304872.7	*	
Cluster 246				
	1	1515746.4	*	
Cluster 247				
	1	1538553.7	*	
Cluster 248				
	1	1543721.5	*	
Cluster 249				
	1	156889.7	*	
Cluster 250				
	1	159087.4	*	
Cluster 251				
	1	194439	*	
Cluster 252				

	1	198628.28	*	
Cluster 253				
	1	243164.3	*	
Cluster 254				
	1	243231	*	
Cluster 255				
	1	262489	*	
Cluster 256				
	1	269799.3	*	
Cluster 257				
	1	272564a4	*	
Cluster 258				
	1	281689	*	
	2	Desulfuromonas		100.00%
Cluster 259				
	1	290317.7	*	
Cluster 260				
	1	290318.4	*	
Cluster 261				
	1	324925.4	*	
Cluster 262				
	1	331678.4	*	
Cluster 263				
	1	3389634	*	
Cluster 264				
	1	340177.8	*	
Cluster 265				
	1	349124.5	*	
Cluster 266				
	1	3516052	*	
	2	Geobacter		100.00%
Cluster 267				
	1	354.5.225	*	
Cluster 268				
	1	354.5.375	*	
Cluster 269				
	1	377629	*	
Cluster 270				
	1	379731.25	*	
Cluster 271				
	1	391774.5	*	
	2	Desulfovibrio		98.47%
Cluster 272				
	1	395493.5	*	
Cluster 273				
	1	398767.12	*	
Cluster 274				
	1	398767.12	*	
Cluster 275				
	1	404380.3	*	
Cluster 276				
	1	404589.4	*	
Cluster 277				
	1	439235.3	*	
Cluster 278				
	1	443143.8	*	
Cluster 279				
	1	447217.4	*	
Cluster 280				
	1	498761.3	*	
Cluster 281				
	1	517417.4	*	
Cluster 282				
	1	517418.3	*	
Cluster 283				
	1	5227723	*	
Cluster 284				
	1	525897.4	*	
Cluster 285				
	1	62928.7	*	

Cluster 286				
	1	1486262.6	*	
	2	HQ456120		98.70%
	3	HQ456103		98.78%
	4	HQ455980		99.69%
	5	HQ455970		100.00%
	6	HQ455958		100.00%
Cluster 287				
	1	6908502	*	
Cluster 288				
	1	697282.9	*	
Cluster 289				
	1	745277.8	*	
Cluster 290				
	1	794903.7	*	
Cluster 291				
	1	883.3	*	
Cluster 292				
	1	1232683.5	*	
Cluster 293				
	1	765909.6	*	
Cluster 294				
	1	6666666.181	*	
Cluster 295				
	1	6666666.181	*	
Cluster 296				
	1	6666666.181	*	
Cluster 297				
	1	6666666.814	*	
Cluster 298				
	1	Erwinia	*	
Cluster 299				
	1	Desulfovibrio	*	
Cluster 300				
	1	Accumulibacter	*	
Cluster 301				
	1	Allochromatium	*	
Cluster 302				
	1	HQ611932	*	
	2	HQ611931		97.55%
	3	HQ611928		97.86%
	4	HQ611865		97.86%
	5	HQ611855		97.86%
	6	HQ611853		97.55%
	7	HQ611835		97.25%
Cluster 303				
	1	HQ611929	*	
	2	HQ611507		97.55%
Cluster 304				
	1	HQ611925	*	
	2	HQ611919		100.00%
	3	HQ611916		100.00%
	4	HQ611879		987.00%
	5	HQ611875		987.00%
	6	HQ611728		97.86%
	7	HQ611725		98.47%
	8	HQ611719		100.00%
	9	HQ611714		100.00%
	10	HQ611448		987.00%
	11	HQ611447		100.00%
	12	HQ611444		98.47%
	13	HQ611442		100.00%
	14	HQ611439		100.00%
	15	HQ611435		98.47%
	16	HQ455861		97.55%
	17	HQ455840		97.86%
	18	EF204561		96.91%
Cluster 305				
	1	HQ611920	*	
Cluster 306				

	1	HQ611911	*	
Cluster 307				
	1	HQ611901	*	
	2	HQ611861		99.08%
Cluster 308				
	1	HQ611892	*	
Cluster 309				
	1	HQ611876	*	
Cluster 310				
	1	HQ611874	*	
Cluster 311				
	1	HQ611850	*	
	2	HQ611735		97.55%
	3	HQ611723		97.55%
	4	HQ611722		97.55%
	5	HQ611557		97.86%
	6	HQ611505		99.08%
	7	HQ611501		98.78%
Cluster 312				
	1	HQ611844	*	
	2	HQ611837		98.47%
Cluster 313				
	1	HQ611791	*	
Cluster 314				
	1	HQ611786	*	
	2	HQ611460		97.86%
Cluster 315				
	1	HQ611782	*	
	2	HQ611741		97.86%
Cluster 316				
	1	HQ611729	*	
	2	HQ611573		99.69%
	3	HQ611567		99.69%
	4	HQ611559		99.69%
Cluster 317				
	1	HQ611642	*	
Cluster 318				
	1	HQ611635	*	
	2	HQ611633		99.69%
	3	HQ611624		99.69%
	4	HQ611622		99.69%
	5	HQ611620		99.69%
	6	HQ611617		99.39%
	7	HQ611615		99.69%
Cluster 319				
	1	HQ611631	*	
Cluster 320				
	1	HQ611623	*	
Cluster 321				
	1	HQ611574	*	
Cluster 322				
	1	HQ611562	*	
Cluster 323				
	1	HQ611542	*	
	2	HQ611534		100.00%
Cluster 324				
	1	HQ611494	*	
Cluster 325				
	1	HQ611438	*	
	2	HQ611433		99.69%
	3	HQ611432		100.00%
	4	HQ611429		100.00%
Cluster 326				
	1	HQ611427	*	
	2	HQ611423		99.69%
	3	HQ611418		99.69%
	4	HQ611416		99.69%
	5	HQ611411		99.39%
Cluster 327				
	1	HQ611425	*	

Cluster 328				
	1	HQ611422	*	
	2	HQ611420		99.69%
Cluster 329				
	1	HQ611419	*	
Cluster 330				
	1	HQ611407	*	
	2	HQ611403		98.47%
Cluster 331				
	1	HQ611389	*	
	2	HQ611388		98.78%
	3	HQ611373		99.39%
Cluster 332				
	1	HQ611383	*	
Cluster 333				
	1	AB727477	*	
Cluster 334				
	1	HQ456093	*	
Cluster 335				
	1	HQ456092	*	
Cluster 336				
	1	HQ456091	*	
	2	HQ456087		99.08%
	3	HQ455849		99.69%
Cluster 337				
	1	HQ456090	*	
Cluster 338				
	1	HQ456089	*	
Cluster 339				
	1	HQ456086	*	
	2	HQ456083		98.78%
	3	HQ456082		99.39%
	4	HQ456081		99.39%
	5	HQ456074		99.39%
Cluster 340				
	1	HQ456085	*	
Cluster 341				
	1	HQ456084	*	
	2	HQ455842		99.39%
Cluster 342				
	1	HQ456080	*	
	2	HQ456077		99.69%
Cluster 343				
	1	HQ456079	*	
Cluster 344				
	1	HQ456078	*	
Cluster 345				
	1	HQ456076	*	
Cluster 346				
	1	HQ456073	*	
Cluster 347				
	1	HQ456072	*	
	2	HQ456070		99.69%
	3	HQ456067		99.69%
	4	HQ456066		100.00%
	5	HQ456065		99.39%
	6	HQ456064		99.39%
	7	HQ456063		99.69%
	8	HQ456062		100.00%
	9	HQ456054		99.39%
	10	HQ456052		99.69%
	11	HQ456048		99.69%
Cluster 348				
	1	HQ456040	*	
	2	HQ456038		99.69%
	3	HQ456034		99.69%
	4	HQ456032		100.00%
	5	HQ456029		100.00%
Cluster 349				
	1	HQ456039	*	

Cluster 350				
	1	HQ456037	*	
	2	EF204563		96.33%
Cluster 351				
	1	HQ456030	*	
Cluster 352				
	1	HQ456020	*	
	2	HQ456011		99.08%
	3	HQ456004		99.69%
Cluster 353				
	1	HQ455993	*	
Cluster 354				
	1	HQ455984	*	
Cluster 355				
	1	HQ455981	*	
	2	HQ455975		99.39%
	3	HQ455973		99.39%
	4	HQ455971		99.39%
Cluster 356				
	1	HQ455977	*	
	2	HQ455974		99.39%
Cluster 357				
	1	HQ455956	*	
	2	HQ455920		99.39%
	3	HQ455919		99.08%
Cluster 358				
	1	HQ455954	*	
	2	HQ455951		99.39%
	3	HQ455946		99.39%
	4	HQ455944		99.08%
	5	HQ455943		99.08%
	6	HQ455934		99.39%
	7	HQ455929		99.39%
	8	HQ455923		99.08%
	9	HQ455921		97.55%
	10	HQ455910		98.78%
	11	HQ455904		99.08%
	12	HQ455893		98.78%
	13	HQ455885		98.78%
Cluster 359				
	1	HQ455950	*	
Cluster 360				
	1	HQ455949	*	
Cluster 361				
	1	HQ455928	*	
Cluster 362				
	1	HQ455914	*	
	2	HQ455907		99.39%
	3	HQ455906		99.69%
	4	HQ455900		99.39%
	5	HQ455892		99.69%
	6	HQ455888		99.39%
	7	HQ455882		99.39%
Cluster 363				
	1	HQ455905	*	
	2	HQ455846		100.00%
Cluster 364				
	1	HQ455903	*	
Cluster 365				
	1	HQ455896	*	
	2	HQ455889		99.39%
Cluster 366				
	1	HQ455884	*	
Cluster 367				
	1	HQ455878	*	
Cluster 368				
	1	HQ455877	*	
	2	GQ475468		97.55%
Cluster 369				
	1	HQ455876	*	

	2	HQ455841		99.69%
Cluster 370				
	1	HQ455874	*	
	2	HQ455850		97.25%
	3	HQ455848		99.69%
Cluster 371				
	1	HQ455873	*	
	2	HQ455863		99.69%
	3	HQ455851		100.00%
Cluster 372				
	1	HQ455872	*	
	2	HQ455871		99.08%
Cluster 373				
	1	HQ455866	*	
Cluster 374				
	1	HQ455862	*	
Cluster 375				
	1	HQ455859	*	
Cluster 376				
	1	HQ455858	*	
	2	HQ455839		99.39%
Cluster 377				
	1	HQ455856	*	
Cluster 378				
	1	HQ455855	*	
	2	HQ455835		99.69%
Cluster 379				
	1	HQ455854	*	
Cluster 380				
	1	HQ455847	*	
Cluster 381				
	1	HQ455845	*	
Cluster 382				
	1	HQ455838	*	
Cluster 383				
	1	EU916538	*	
	2	EU916536		100.00%
Cluster 384				
	1	EU916420	*	
	2	EU916418		98.78%
	3	EU916412		98.78%
	4	EU916411		98.78%
	5	EU916407		99.39%
	6	EU916404		99.39%
	7	EU916390		99.39%
	8	EU916384		99.08%
Cluster 385				
	1	EU916414	*	
	2	EU916405		98.78%
	3	EU916394		99.39%
	4	EU916391		99.08%
	5	EU916387		98.78%
Cluster 386				
	1	EU916382	*	
Cluster 387				
	1	EU916330	*	
	2	EU916322		99.69%
	3	EU916321		99.39%
Cluster 388				
	1	EU916320	*	
	2	EU916296		99.08%
Cluster 389				
	1	EU916319	*	
Cluster 390				
	1	EU916318	*	
	2	EU916315		98.47%
	3	EU916314		99.39%
	4	EU916313		99.69%
	5	EU916312		99.39%
	6	EU916310		98.78%

	7	EU916309	99.69%
	8	EU916305	99.08%
	9	EU916302	99.08%
	10	EU916300	99.69%
	11	EU916299	99.39%
	12	EU916298	99.69%
	13	EU916294	99.39%
	14	EU916293	99.69%
	15	EU916291	99.39%
	16	EU916290	99.39%
	17	EU916289	99.69%
	18	EU916282	99.69%
	19	EU916281	99.39%
	20	EU916279	99.69%
	21	EU916278	99.39%
	22	EU916277	99.39%
	23	EU916276	99.69%
	24	EU916275	99.08%
Cluster 391			
	1	KC140473	*
	2	KC140467	100.00%
	3	KC140463	100.00%
	4	KC140459	99.69%
	5	KC140456	100.00%
	6	KC140454	99.39%
	7	KC140453	99.69%
	8	KC140447	100.00%
	9	KC140444	100.00%
	10	KC140441	99.39%
	11	KC140439	100.00%
	12	KC140435	99.69%
	13	KC140433	100.00%
	14	KC140428	99.69%
	15	KC140427	100.00%
	16	KC140473	100.00%
	17	KC140467	100.00%
	18	KC140463	100.00%
	19	KC140459	99.69%
	20	KC140456	100.00%
	21	KC140454	99.39%
	22	KC140453	99.69%
	23	KC140447	100.00%
	24	KC140444	100.00%
	25	KC140441	99.39%
	26	KC140439	100.00%
	27	KC140435	99.69%
	28	KC140433	100.00%
	29	KC140428	99.69%
	30	KC140427	100.00%
Cluster 392			
	1	KC140438	*
	2	KC140438	100.00%
Cluster 393			
	1	KC140424	*
	2	KC140424	100.00%
Cluster 394			
	1	GQ475479	*
	2	GQ475476	99.08%
Cluster 395			
	1	GQ475475	*
Cluster 396			
	1	GQ475474	*
Cluster 397			
	1	GQ475473	*
	2	GQ475470	99.39%
	3	GQ475469	99.69%
	4	GQ475466	99.39%
	5	GQ475462	99.69%
	6	GQ475459	98.78%
	7	GQ475458	99.08%

	8	GQ475457		96.33%
	9	GQ475456		99.69%
	10	GQ475441		98.70%
	11	GQ475437		99.08%
	12	GQ475430		97.86%
	13	GQ475429		97.25%
Cluster 398				
	1	GQ475472	*	
Cluster 399				
	1	GQ475471	*	
Cluster 400				
	1	GQ475465	*	
	2	GQ475464		98.47%
	3	GQ475463		98.70%
	4	GQ475461		97.25%
	5	GQ475460		96.94%
Cluster 401				
	1	GQ475455	*	
Cluster 402				
	1	GQ475454	*	
Cluster 403				
	1	GQ475453	*	
Cluster 404				
	1	GQ475452	*	
Cluster 405				
	1	GQ475451	*	
Cluster 406				
	1	GQ475449	*	
	2	GQ475447		99.08%
Cluster 407				
	1	GQ475444	*	
Cluster 408				
	1	GQ475442	*	
Cluster 409				
	1	GQ475440	*	
Cluster 410				
	1	GQ475439	*	
Cluster 411				
	1	GQ475438	*	
Cluster 412				
	1	GQ475436	*	
Cluster 413				
	1	AY896432	*	
Cluster 414				
	1	AY896430	*	
	2	AY896429		99.39%
	3	AY896428		99.69%
Cluster 415				
	1	EF204557	*	
Cluster 416				
	1	AY896469	*	
	2	KF151808		99.07%
	3	DQ481395		98.50%
	4	DQ481393		98.46%
	5	DQ481392		97.84%
	6	DQ481384		98.50%
	7	DQ481383		96.91%
	8	DQ481370		97.53%
	9	HQ229034		99.38%
	10	HQ229032		99.07%
	11	HQ229025		98.77%
	12	HQ229021		99.07%
	13	HQ229017		98.77%
	14	KC013201		98.77%
	15	KC013174		98.46%
	16	KC013171		98.46%
	17	KC013166		99.07%
	18	KC013164		99.07%
	19	KC013163		98.46%
	20	KC013160		98.46%

21	KC013157	99.07%
22	KC013154	99.07%
23	KC013153	98.46%
24	KC013152	98.77%
25	KC013151	98.77%
26	KC013148	99.07%
27	KC013146	99.07%
28	KC013144	98.46%
29	KC013141	98.77%
30	KC013139	98.77%
31	KC013137	99.07%
32	KC013136	99.07%
33	KC013134	99.07%
34	KC013133	98.77%
35	KC013132	98.77%
36	KC013131	99.07%
37	KC013129	99.07%
38	KC013128	98.46%
39	KC013127	98.77%
40	KC013126	99.07%
41	KC013124	99.07%
42	KC013123	98.77%
43	KC013119	99.07%
44	KC013118	99.07%
45	KC013116	99.07%
46	KC013115	98.77%
47	KC013114	99.07%
48	KC013112	99.07%
49	KC013110	99.07%
50	KC013107	99.07%
51	KC013106	98.77%
52	KC013102	98.46%
53	KC013101	99.07%
54	KC013100	99.07%
55	KC013099	99.07%
56	KC013097	99.07%
57	KC013096	98.77%
58	KC013094	98.46%
59	KC013093	98.77%
60	KC013090	99.07%
61	KC013086	98.50%
62	KC013085	98.46%
63	KC013084	98.77%
64	KC013083	99.07%
65	KC013077	99.07%
66	KC013073	99.07%
67	KC013072	99.07%
68	KC013071	99.07%
69	KC013067	99.07%
70	KC013065	99.07%
71	KC013064	98.46%
72	KC013063	99.07%
73	KC013062	99.07%
74	KC013061	99.07%
75	KC013060	98.46%
76	KC013056	98.50%
77	KC013055	99.07%
78	KC013046	99.07%
79	KC013041	98.77%
80	EU052396	99.07%
81	DQ825728	99.07%
82	KF151809	98.77%
83	KF151807	98.77%
84	HQ660931	98.46%
85	HQ660847	98.77%
86	HQ660840	98.77%
87	HQ660837	99.07%
88	HQ660834	99.07%
89	HQ660830	99.07%
90	HQ660828	98.77%

91	HQ660827	99.07%
92	AY191962.2	98.50%
93	Candidatus	99.38%
94	HQ611827	99.07%
95	HQ611826	99.07%
96	HQ611823	99.07%
97	HQ611822	99.38%
98	HQ611815	99.38%
99	HQ611814	98.50%
100	HQ611811	99.38%
101	HQ611807	99.38%
102	HQ611805	99.07%
103	HQ611799	99.07%
104	HQ611796	99.07%
105	HQ611789	98.77%
106	HQ611779	98.77%
107	HQ611776	99.07%
108	HQ611775	98.77%
109	HQ611772	99.38%
110	HQ611767	99.07%
111	HQ611753	99.07%
112	HQ611751	99.07%
113	HQ611749	99.07%
114	HQ611734	99.07%
115	HQ611726	99.07%
116	HQ611721	99.07%
117	HQ611709	99.07%
118	HQ611662	99.07%
119	HQ611652	99.07%
120	HQ611650	98.77%
121	HQ611648	99.38%
122	HQ611647	99.07%
123	HQ611639	99.38%
124	HQ611613	99.38%
125	HQ611607	99.38%
126	HQ611594	99.07%
127	HQ611589	99.38%
128	HQ611587	99.38%
129	HQ611586	98.46%
130	HQ611580	99.38%
131	HQ611579	99.07%
132	HQ611577	99.38%
133	HQ611370	99.07%
134	HQ611366	99.38%
135	HQ611363	99.38%
136	AB727505	99.07%
137	AB727504	99.38%
138	AB727496	99.38%
139	AB727493	99.07%
140	AB727492	99.38%
141	HQ456071	99.07%
142	HQ456069	98.77%
143	HQ456068	98.50%
144	HQ456060	99.38%
145	HQ456059	98.46%
146	HQ456057	98.77%
147	HQ456053	99.07%
148	HQ456051	98.77%
149	HQ456049	99.07%
150	AY896468	99.08%
151	AY896467	99.08%
152	AY896466	99.69%
153	AY896465	99.08%
154	AY896464	98.77%
155	AY896463	98.77%
156	AY896440	99.08%
157	AY896439	98.77%
158	AY896438	99.39%
159	AY896437	99.07%
160	AY896436	98.77%

161	AY896435		98.77%
162	AY896434		98.46%
163	AY896433		98.77%
164	AY896427		98.50%
165	AY896426		98.77%
166	AY896425		99.07%
167	AY896424		98.77%
168	AY896422		99.07%
169	AY896421		99.07%
170	AY896420		99.07%
171	AY896419		98.50%
172	AY896418		99.07%
173	AY896417		99.07%
174	AY896361		99.08%
175	AY896360		99.38%
176	AY896359		99.08%
177	AY896358		98.77%
178	AY896357		98.47%
179	AY896356		98.77%
180	AY896355		99.08%
181	AY896354		98.60%
Cluster 417			
1	AY896454	*	
2	AF536986		98.77%
3	AF536985		98.77%
4	AF536984		98.50%
5	AF536983		98.50%
6	DQ481415		98.77%
7	DQ481414		99.38%
8	DQ481412		99.07%
9	DQ481410		98.46%
10	DQ481409		98.77%
11	DQ481408		97.84%
12	DQ481407		99.07%
13	DQ481406		98.77%
14	DQ481405		98.77%
15	DQ481404		98.77%
16	DQ481403		98.50%
17	DQ481402		99.07%
18	DQ481401		98.77%
19	DQ481399		99.07%
20	DQ481341		98.77%
21	DQ481340		97.22%
22	DQ481339		99.07%
23	DQ481338		98.77%
24	DQ481337		98.77%
25	DQ481241		99.07%
26	HQ229018		98.77%
27	KC013229		99.07%
28	KC013228		99.07%
29	KC013227		99.07%
30	KC013221		98.46%
31	AB928305		99.69%
32	Crocosphaera		99.38%
33	HQ611913		99.38%
34	HQ456119		98.77%
35	HQ456027		99.07%
36	HQ456021		99.07%
37	HQ456018		99.38%
38	HQ456003		99.38%
39	HQ456002		98.46%
40	HQ455997		99.38%
41	HQ455987		98.77%
42	HQ455986		99.07%
43	HQ455899		99.38%
44	AY896384		98.77%
Cluster 418			
1	AY896414	*	
Cluster 419			
1	AY896410	*	

	2	AY896336		96.01%
	3	AY896322		97.24%
Cluster 420				
	1	AY896377	*	
Cluster 421				
	1	AY896342	*	
	2	KC013178		98.77%
	3	KC013170		99.07%
	4	HQ456118		99.07%
	5	HQ456113		99.07%
	6	HQ456100		99.07%
	7	HQ456097		98.77%
	8	HQ456058		98.46%
	9	HQ456056		98.46%
	10	HQ456055		98.50%
	11	HQ456050		98.50%
	12	HQ456022		98.77%
	13	HQ455982		97.84%
	14	HQ455972		99.07%
	15	HQ455969		98.46%
	16	HQ455927		98.46%
	17	HQ455916		99.07%
	18	HQ455912		99.07%
	19	HQ455886		98.46%
	20	AY896461		99.08%
	21	AY896343		98.77%
	22	AY896341		96.63%
	23	AY896340		98.60%
Cluster 422				
	1	AY896310	*	
Cluster 423				
	1	HM801405	*	
	2	JF429966		97.84%
	3	JF429964		99.38%
	4	DQ481398		96.91%
	5	DQ481397		97.22%
	6	DQ481396		96.91%
	7	DQ481388		96.30%
	8	DQ481387		96.91%
	9	DQ481385		96.60%
	10	DQ481378		96.60%
	11	DQ481311		96.91%
	12	DQ481309		96.60%
	13	DQ481307		96.30%
	14	DQ481304		96.91%
	15	DQ481303		96.60%
	16	DQ481302		96.60%
	17	DQ481301		96.60%
	18	DQ481299		96.30%
	19	DQ481297		96.60%
	20	DQ481295		96.91%
	21	DQ481292		96.91%
	22	DQ481291		96.91%
	23	DQ481290		96.91%
	24	DQ481257		96.91%
	25	DQ481251		96.91%
	26	DQ481248		96.91%
	27	DQ481247		96.30%
	28	DQ481246		96.91%
	29	DQ481244		96.91%
	30	DQ481242		96.91%
	31	EU052520		96.91%
	32	EU151800		96.91%
	33	EU151799		96.91%
	34	EU151798		96.60%
	35	EF568528		99.07%
	36	EF568524		98.77%
	37	EF568519		99.07%
	38	EF568505		99.07%
	39	EF568503		99.07%

	40	EF568501	99.07%
	41	EF568498	99.07%
	42	EF568496	99.07%
	43	EF568495	99.07%
	44	HM801409	99.69%
	45	HM801399	99.69%
	46	HM801385	99.69%
	47	HM801383	99.69%
	48	HM801380	99.69%
	49	HM801379	99.38%
	50	HM801377	99.69%
	51	HM801375	99.69%
	52	HM801374	99.69%
	53	HM801373	99.69%
	54	HM801368	99.69%
	55	1265504.4	99.07%
	56	HQ611702	98.46%
	57	KC140383	98.46%
	58	KC140374	97.84%
	59	KC140373	98.77%
	60	KC140372	98.77%
	61	KC140371	98.77%
	62	KC140370	97.84%
	63	KC140369	98.77%
	64	KC140471	98.46%
	65	KC140470	98.77%
	66	KC140466	98.50%
	67	KC140465	98.46%
	68	KC140464	98.77%
	69	KC140458	98.77%
	70	KC140457	98.46%
	71	KC140451	98.77%
	72	KC140450	98.46%
	73	KC140443	98.77%
	74	KC140437	98.46%
	75	KC140436	98.50%
	76	KC140431	98.77%
	77	KC140430	98.50%
	78	KC140429	98.77%
	79	KC140423	98.77%
	80	KC140422	97.22%
	81	KC140383	98.46%
	82	KC140374	97.84%
	83	KC140373	98.77%
	84	KC140372	98.77%
	85	KC140371	98.77%
	86	KC140370	97.84%
	87	KC140369	98.77%
	88	KC140471	98.46%
	89	KC140470	98.77%
	90	KC140466	98.50%
	91	KC140465	98.46%
	92	KC140464	98.77%
	93	KC140458	98.77%
	94	KC140457	98.46%
	95	KC140451	98.77%
	96	KC140450	98.46%
	97	KC140443	98.77%
	98	KC140437	98.46%
	99	KC140436	98.50%
	100	KC140431	98.77%
	101	KC140430	98.50%
	102	KC140429	98.77%
	103	KC140423	98.77%
	104	KC140422	97.22%
Cluster 424			
	1	HM801365	*
	2	HM801371	99.38%
	3	HM801369	99.38%
	4	HM801361	99.38%

Cluster 425

1	HM801221	*	
2	HM801597		98.77%
3	HM801477		98.46%
4	HM801476		98.46%
5	HM801475		97.84%
6	HM801474		98.46%
7	HM801473		99.69%
8	HM801472		98.46%
9	HM801469		98.46%
10	HM801466		98.46%
11	HM801465		98.46%
12	HM801332		98.46%
13	HM801331		98.46%
14	HM801330		98.50%
15	HM801328		98.46%
16	HM801327		98.50%
17	HM801326		98.46%
18	HM801325		98.46%
19	HM801323		98.50%
20	HM801322		98.50%
21	HM801321		98.46%
22	HM801320		97.84%
23	HM801319		98.46%
24	HM801317		98.46%
25	HM801316		98.46%
26	HM801314		98.46%
27	HM801313		98.46%
28	HM801312		97.84%
29	HM801311		98.77%
30	HM801309		98.46%
31	HM801308		98.46%
32	HM801307		98.46%
33	HM801302		98.46%
34	HM801300		98.46%
35	HM801296		98.46%
36	HM801294		98.50%
37	HM801293		98.46%
38	HM801291		98.46%
39	HM801290		98.46%
40	HM801287		98.46%
41	HM801286		98.50%
42	HM801285		98.46%
43	HM801281		98.46%
44	HM801278		98.46%
45	HM801275		98.46%
46	HM801274		98.50%
47	HM801273		98.46%
48	HM801270		98.46%
49	HM801269		98.46%
50	HM801266		98.46%
51	HM801256		98.50%
52	HM801252		98.46%
53	HM801241		98.46%
54	HM801240		97.84%
55	HM801232		98.46%
56	HM801231		98.46%
57	HM801224		98.50%
58	HM801223		98.46%
59	HM801220		98.46%
60	HM801219		98.46%
61	HM801214		98.46%
62	HM801207		98.46%
63	HM801204		98.46%
64	HM801200		98.50%
65	HM801198		98.46%
66	HM801195		98.46%
67	HM801192		98.46%
68	HM801187		98.46%
69	HM801185		98.46%

	70	HM801184	97.84%
	71	HM801182	98.46%
	72	HM801181	98.46%
	73	HM801180	98.46%
Cluster 426			
	1	HM801595	*
	2	HM801791	97.53%
	3	JF429952	98.46%
	4	HM801836	97.84%
	5	HM801835	96.91%
	6	HM801833	98.46%
	7	HM801831	98.50%
	8	HM801829	98.46%
	9	HM801824	98.46%
	10	HM801821	96.91%
	11	HM801820	98.46%
	12	HM801819	98.46%
	13	HM801817	98.50%
	14	HM801812	97.22%
	15	HM801811	98.46%
	16	HM801809	97.84%
	17	HM801806	97.22%
	18	HM801803	97.53%
	19	HM801795	98.46%
	20	HM801793	98.46%
	21	HM801779	98.50%
	22	HM801778	98.46%
	23	HM801776	98.46%
	24	HM801766	98.77%
	25	HM801750	98.46%
	26	HM801742	98.46%
	27	HM801734	97.22%
	28	HM801732	98.50%
	29	HM801730	98.46%
	30	HM801728	97.22%
	31	HM801727	98.46%
	32	HM801726	97.22%
	33	HM801719	98.50%
	34	HM801716	97.22%
	35	HM801714	97.22%
	36	HM801710	98.46%
	37	HM801707	98.46%
	38	HM801706	98.46%
	39	HM801705	97.53%
	40	HM801697	98.46%
	41	HM801691	97.22%
	42	HM801688	97.22%
	43	HM801686	98.46%
	44	HM801682	96.91%
	45	HM801677	97.84%
	46	HM801676	98.46%
	47	HM801675	98.46%
	48	HM801672	98.46%
	49	HM801669	98.46%
	50	HM801668	98.46%
	51	HM801667	98.46%
	52	HM801665	98.46%
	53	HM801664	98.50%
	54	HM801663	98.46%
	55	HM801661	97.84%
	56	HM801658	98.46%
	57	HM801656	98.46%
	58	HM801653	98.46%
	59	HM801652	98.50%
	60	HM801650	98.77%
	61	HM801646	96.91%
	62	HM801644	97.22%
	63	HM801643	98.50%
	64	HM801640	98.46%
	65	HM801634	98.50%

	66	HM801632		97.22%
	67	HM801624		98.46%
	68	HM801622		98.50%
	69	HM801620		98.46%
	70	HM801617		98.46%
	71	HM801614		97.84%
	72	HM801613		98.50%
	73	HM801606		98.46%
	74	HM801602		99.07%
Cluster 427				
	1	HM801229	*	
Cluster 428				
	1	HM801441	*	
	2	HM801401		98.50%
Cluster 429				
	1	HM801494	*	
	2	HM801513		96.91%
	3	HM801502		96.91%
	4	HM801482		96.91%
	5	HM801481		96.91%
	6	HM801480		96.91%
	7	HM801458		97.22%
	8	HM801449		96.91%
	9	HM801419		96.91%
	10	HM801414		96.91%
	11	HM801413		96.60%
	12	HM801412		97.22%
Cluster 430				
	1	HM801504	*	
	2	HM801512		99.38%
	3	HM801511		99.69%
	4	HM801509		99.07%
	5	HM801508		99.07%
	6	HM801501		99.07%
	7	HM801495		99.07%
	8	HM801489		98.46%
	9	HM801486		99.07%
	10	HM801408		99.07%
	11	HM801400		99.07%
	12	HM801394		99.07%
	13	HM801391		99.07%
	14	HM801355		99.07%
	15	HM801352		99.07%
	16	HM801350		98.46%
	17	HM801348		99.07%
	18	HM801343		98.77%
	19	HM801341		98.77%
	20	HM801340		99.07%
	21	HM801339		98.77%
	22	HM801338		99.07%
	23	HM801334		98.77%
Cluster 431				
	1	HM801565	*	
	2	AB928264		99.69%
	3	AB928261		99.69%
	4	HM801558		99.38%
	5	HM801557		98.77%
	6	HM801556		98.50%
	7	HM801552		98.46%
	8	HM801528		98.46%
	9	HM801517		99.69%
	10	HM801282		99.69%
Cluster 432				
	1	HM801493	*	
	2	KC013142		99.38%
	3	KC013122		99.07%
	4	KC013043		99.07%
	5	HM801510		99.07%
	6	HM801500		98.77%
	7	HM801488		99.07%

	8	HM801485	99.07%
	9	HM801483	98.77%
	10	HM801478	99.07%
Cluster 433			
	1	HM801648	*
	2	HM801638	99.38%
	3	HM801631	99.07%
Cluster 434			
	1	KF151422	*
	2	KF151447	98.50%
	3	KF151421	98.46%
	4	KF151418	98.46%
Cluster 435			
	1	KF151439	*
	2	DQ481372	97.22%
	3	DQ481368	96.60%
	4	DQ481366	96.60%
	5	DQ481361	96.91%
	6	DQ481360	96.91%
	7	DQ481359	96.60%
	8	DQ481358	97.22%
	9	DQ481262	96.91%
	10	DQ481255	96.60%
	11	DQ481254	96.60%
	12	DQ481253	96.60%
	13	DQ481357	96.30%
	14	DQ481356	96.91%
	15	DQ481355	96.60%
	16	DQ481349	96.60%
	17	DQ481347	96.30%
	18	DQ481346	96.60%
	19	DQ481344	96.30%
	20	KF151444	98.77%
	21	KF151443	99.69%
	22	KF151442	98.77%
	23	KF151441	98.77%
	24	KF151440	98.46%
	25	KF151438	99.07%
	26	KF151436	98.77%
	27	KF151435	98.77%
	28	KF151434	98.77%
	29	KF151433	98.77%
	30	KF151432	98.46%
	31	KF151431	98.46%
	32	KF151428	99.07%
Cluster 436			
	1	KF151417	*
	2	KF151420	99.69%
	3	KF151414	99.69%
Cluster 437			
	1	HQ660913	*
Cluster 438			
	1	JF429973	*
	2	JF429972	99.69%
	3	JF429971	98.50%
	4	KC013234	99.38%
	5	KC013231	99.38%
	6	KC013219	99.07%
	7	KC013218	99.38%
	8	KC013216	98.50%
	9	KC013215	98.50%
	10	KC013214	99.07%
	11	KC013211	99.07%
	12	KC013209	98.77%
	13	KC013207	98.50%
	14	KC013205	99.07%
	15	KC013204	99.07%
	16	KC013203	99.07%
	17	KC013200	98.77%
	18	KC013197	98.77%

	19	HQ611487		99.69%
	20	HQ611478		99.38%
	21	HQ611477		99.38%
	22	HQ611476		99.69%
	23	HQ611470		99.69%
	24	HQ611467		99.69%
Cluster 439				
	1	JF429963	*	
Cluster 440				
	1	JF429962	*	
	2	JF429959		96.91%
	3	JF429956		99.38%
	4	JF429954		96.60%
	5	JF429953		96.30%
	6	JF429951		96.60%
	7	JF429950		96.60%
	8	JF429948		96.60%
	9	JF429947		96.60%
	10	JF429945		96.60%
	11	JF429944		96.30%
	12	EU052420		98.46%
	13	EU052418		98.46%
	14	EU052417		98.77%
	15	EU052416		97.84%
	16	EU052404		98.50%
	17	EU052346		98.46%
	18	AB928303		96.60%
	19	AB928302		96.30%
	20	AB928299		96.30%
	21	AB928295		96.30%
	22	AB928294		96.30%
	23	HM801840		96.60%
	24	HM801839		96.30%
	25	HM801837		96.60%
	26	HM801834		96.30%
	27	HM801832		96.30%
	28	HM801828		96.30%
	29	HM801827		96.30%
	30	HM801826		96.30%
	31	HM801825		96.30%
	32	HM801823		96.30%
	33	HM801822		96.30%
	34	HM801818		96.30%
	35	HM801816		96.30%
	36	HM801815		96.30%
	37	HM801814		96.60%
	38	HM801813		96.30%
	39	HM801810		96.30%
	40	HM801808		96.30%
	41	HM801807		96.30%
	42	HM801805		96.60%
	43	HM801804		96.30%
	44	HM801802		96.60%
	45	HM801801		96.60%
	46	HM801800		96.91%
	47	HM801799		96.30%
	48	HM801798		96.30%
	49	HM801797		96.30%
	50	HM801796		96.30%
	51	HM801794		96.91%
	52	HM801792		96.30%
	53	HM801790		96.60%
	54	HM801789		96.60%
	55	HM801788		96.60%
	56	HM801787		96.60%
	57	HM801786		96.30%
	58	HM801785		96.60%
	59	HM801784		96.60%
	60	HM801783		96.30%
	61	HM801782		96.30%

62	HM801781	96.30%
63	HM801777	96.30%
64	HM801775	96.30%
65	HM801774	96.30%
66	HM801773	96.30%
67	HM801772	96.30%
68	HM801768	96.30%
69	HM801760	96.30%
70	HM801753	96.30%
71	HM801736	96.30%
72	HM801735	96.30%
73	HM801733	96.30%
74	HM801731	96.30%
75	HM801729	96.30%
76	HM801725	96.30%
77	HM801724	96.60%
78	HM801723	96.30%
79	HM801722	96.30%
80	HM801721	96.60%
81	HM801720	96.30%
82	HM801718	96.30%
83	HM801717	96.60%
84	HM801715	96.60%
85	HM801713	96.30%
86	HM801712	96.30%
87	HM801711	97.22%
88	HM801709	96.60%
89	HM801704	96.60%
90	HM801703	96.60%
91	HM801702	96.30%
92	HM801701	96.60%
93	HM801700	96.60%
94	HM801699	96.60%
95	HM801698	96.60%
96	HM801696	96.30%
97	HM801695	96.30%
98	HM801694	96.30%
99	HM801693	96.60%
100	HM801692	96.30%
101	HM801690	96.30%
102	HM801689	96.30%
103	HM801687	96.30%
104	HM801685	96.60%
105	HM801684	96.60%
106	HM801683	96.60%
107	HM801681	96.60%
108	HM801680	96.60%
109	HM801679	96.30%
110	HM801678	96.30%
111	HM801674	96.30%
112	HM801673	96.60%
113	HM801671	96.30%
114	HM801670	96.30%
115	HM801666	96.30%
116	HM801662	96.30%
117	HM801660	96.30%
118	HM801657	96.30%
119	HM801655	96.30%
120	HM801654	96.30%
121	HM801651	96.30%
122	HM801649	96.30%
123	HM801647	96.60%
124	HM801645	96.60%
125	HM801642	96.60%
126	HM801641	96.30%
127	HM801637	96.30%
128	HM801636	96.60%
129	HM801635	97.22%
130	HM801633	96.91%
131	HM801630	96.60%

	132	HM801629	96.30%
	133	HM801628	96.30%
	134	HM801627	96.60%
	135	HM801626	96.30%
	136	HM801625	96.30%
	137	HM801621	96.30%
	138	HM801619	96.30%
	139	HM801616	96.30%
	140	HM801615	96.91%
	141	HM801612	96.60%
	142	HM801611	96.30%
	143	HM801609	96.30%
	144	HM801608	96.30%
	145	HM801607	96.30%
	146	HM801605	96.30%
	147	HM801604	96.30%
	148	HM801603	96.30%
	149	HM801601	96.91%
	150	HM801600	96.60%
	151	HM801598	96.60%
	152	HM801431	96.30%
	153	HM801428	96.30%
	154	HM801237	96.30%
	155	HM801230	96.30%
	156	KF151477	96.30%
	157	6666666b180340.pe	98.46%
	158	HQ455955	97.53%
	159	HQ455941	97.53%
	160	HQ455938	97.22%
Cluster 441			
	1	JF429960	*
	2	EF568482	97.84%
	3	EF568481	985.00%
	4	EF568477	985.00%
Cluster 442			
	1	JF429949	*
	2	Bradyrhizobium	97.53%
	3	Bradyrhizobium	98.50%
	4	HQ611531	98.50%
	5	HQ611529	98.46%
	6	HQ611526	98.46%
	7	HQ611522	98.50%
	8	HQ611518	98.46%
Cluster 443			
	1	JF429943	*
Cluster 444			
	1	KF151413	*
Cluster 445			
	1	DQ481394	*
	2	KJ494879	98.77%
Cluster 446			
	1	DQ481389	*
	2	DQ481377	97.22%
	3	DQ481310	96.91%
	4	DQ481261	96.91%
	5	DQ481252	97.53%
	6	EU052518	97.22%
	7	EU151795	96.91%
Cluster 447			
	1	DQ481381	*
	2	DQ481379	99.07%
	3	DQ481376	99.07%
	4	DQ481374	99.38%
Cluster 448			
	1	DQ481324	*
	2	DQ481318	96.91%
	3	HQ611939	96.30%
	4	HQ611905	96.30%
	5	AY896351	96.28%
Cluster 449			

	1	DQ481308	*	
Cluster 450				
	1	DQ481294	*	
Cluster 451				
	1	DQ481258	*	
	2	DQ481352		97.22%
	3	DQ481345		96.60%
Cluster 452				
	1	HQ229035	*	
	2	HQ229028		99.07%
	3	HQ229026		98.77%
	4	HQ229008		99.38%
	5	HQ229007		99.07%
	6	HQ229006		98.50%
	7	DQ825746		98.77%
	8	DQ825741		99.07%
	9	DQ825734		98.77%
	10	DQ825716		99.07%
	11	DQ825715		99.07%
	12	HQ611387		98.50%
	13	HQ611386		98.50%
	14	HQ611384		99.07%
	15	HQ611382		98.77%
	16	HQ611381		99.07%
	17	HQ611379		99.38%
	18	HQ611375		99.07%
	19	HQ611374		98.46%
	20	HQ611372		98.77%
	21	EU916466		99.38%
	22	EU916325		99.38%
	23	EU916284		99.38%
Cluster 453				
	1	HQ229033	*	
Cluster 454				
	1	HQ229014	*	
	2	HQ229011		99.69%
	3	HQ611843		97.53%
	4	HQ611832		97.53%
	5	HQ611682		97.53%
	6	HQ455894		97.53%
Cluster 455				
	1	HQ229012	*	
	2	HQ455926		96.91%
Cluster 456				
	1	HM210402	*	
	2	HM210401		98.46%
	3	HM210400		97.22%
	4	HM210399		98.46%
	5	HM210392		98.46%
	6	HM210391		97.84%
	7	HM210389		97.22%
	8	HM210387		98.50%
Cluster 457				
	1	HM210398	*	
Cluster 458				
	1	KC013224	*	
Cluster 459				
	1	KC013196	*	
	2	KC013191		99.69%
	3	KC013190		99.69%
	4	KC013187		99.69%
	5	KC013181		99.69%
Cluster 460				
	1	DQ831008	*	
Cluster 461				
	1	DQ831007	*	
	2	DQ825749		97.84%
Cluster 462				
	1	DQ825753	*	
Cluster 463				

	1	DQ825745	*
Cluster 464			
	1	DQ825738	*
	2	DQ825736	99.07%
Cluster 465			
	1	DQ825733	*
Cluster 466			
	1	DQ825720	*
	2	DQ825717	99.69%
Cluster 467			
	1	EF568561	*
	2	EF568557	99.38%
Cluster 468			
	1	EF568516	*
	2	EF568514	99.38%
	3	EF568510	99.69%
Cluster 469			
	1	EF568492	*
	2	EF568491	99.07%
	3	EF568490	99.38%
	4	EF568488	99.07%
	5	EF568487	98.77%
Cluster 470			
	1	EF568447	*
	2	EF568444	99.69%
	3	EF568465	99.07%
	4	EF568463	99.69%
	5	EF568462	99.38%
	6	EF568461	99.69%
	7	EF568460	99.69%
	8	EF568456	99.69%
	9	EF568453	99.38%
	10	EF568450	98.77%
Cluster 471			
	1	EF568565	*
	2	EF568542	99.69%
	2	KJ494878	99.38%
	3	KJ494886	99.69%
	4	HM042893	99.38%
	5	HM042892	99.07%
	6	HM042881	98.46%
	7	HM042880	98.77%
	8	Contaminant	99.07%
	9	KF151461	99.38%
	10	KF151460	99.07%
	11	KF151458	99.38%
	12	EU916514	97.53%
	13	EU916498	99.38%
	14	EU916480	99.69%
	15	EU916469	99.38%
	16	EU916706	99.69%
	17	EU916705	98.46%
	18	EU916704	99.69%
	19	EU916701	99.38%
	20	EU916700	99.69%
	21	EU916698	99.07%
	22	EU916697	99.69%
	23	EU916696	99.38%
	24	EU916695	99.38%
	25	EU916694	99.07%
	26	EU916693	99.69%
	27	EU916692	99.38%
	28	EU916691	99.69%
	29	EU916690	99.38%
	30	EU916689	99.69%
	31	EU916687	99.38%
	32	EU916686	99.38%
	33	EU916685	98.77%
	34	EU916684	99.69%
	35	EU916683	99.07%

	36	EU916682		99.69%
	37	EU916680		99.38%
	38	EU916679		99.38%
	39	EU916678		99.69%
	40	EU916676		99.38%
	41	EU916675		99.38%
	42	EU916674		99.69%
	43	EU916671		99.07%
	44	EU916669		99.07%
	45	EU916668		99.07%
	46	EU916666		99.07%
	47	EU916665		99.69%
	48	EU916664		99.38%
	49	EU916663		99.69%
Cluster 472				
	1	EF568554	*	
	2	EF568553		99.07%
	3	EF568552		99.69%
Cluster 473				
	1	EF568476	*	
Cluster 474				
	1	KJ494876	*	
	2	KJ494883		98.77%
	3	KJ494888		97.84%
Cluster 475				
	1	KJ494877	*	
	2	KJ494882		97.22%
	3	KJ494884		97.53%
Cluster 476				
	1	KJ494880	*	
Cluster 477				
	1	KJ494881	*	
Cluster 478				
	1	Richelia sp.	*	
Cluster 479				
	1	Desulfobacter curvatus	*	
Cluster 480				
	1	Paenibacillus azotofixans	*	
Cluster 481				
	1	Paenibacillus azotofixans	*	
Cluster 482				
	1	Frankia sp.	*	
	2	298653.4		99.69%
Cluster 483				
	0	Xanthobacter flavus	*	
	1	1131814.4		98.46%
Cluster 484				
	0	Methanothermobacter	*	
Cluster 485				
	0	Methanococcus vannielii	*	
Cluster 486				
	1	Contaminant	*	
	2	EU916607		99.69%
	3	EU916604		99.38%
	4	EU916602		99.69%
	5	EU916600		99.38%
	6	EU916595		99.38%
	7	EU916594		99.38%
	8	EU916592		99.69%
	9	EU916589		99.69%
	10	EU916585		99.69%
	11	EU916580		99.07%
	12	EU916578		99.07%
	13	EU916577		99.38%
	14	EU916576		99.69%
	15	EU916574		99.38%
	16	EU916573		99.38%
	17	EU916572		99.69%
	18	EU916570		99.38%
	19	EU916569		99.69%

20	EU916566	99.69%
21	EU916565	99.38%
22	EU916563	99.69%
23	EU916561	99.38%
24	EU916560	99.38%
25	EU916558	99.69%
26	EU916554	98.77%
27	EU916553	99.69%
28	EU916551	99.38%
29	EU916545	99.38%
30	EU916544	99.69%
31	EU916543	99.38%
32	EU916542	99.07%
33	EU916540	99.69%
34	EU916537	99.38%
35	EU916530	99.38%
36	EU916528	99.38%
37	EU916526	99.69%
38	EU916522	99.69%
39	EU916519	98.77%
40	EU916518	99.69%
41	EU916515	99.38%
42	EU916513	99.38%
43	EU916512	99.69%
44	EU916511	99.07%
45	EU916510	99.07%
46	EU916509	99.38%
47	EU916508	99.69%
48	EU916507	99.07%
49	EU916505	99.38%
50	EU916503	99.69%
51	EU916502	98.46%
52	EU916501	99.69%
53	EU916496	99.69%
54	EU916491	99.07%
55	EU916488	99.69%
56	EU916486	99.38%
57	EU916485	99.69%
58	EU916484	98.46%
59	EU916483	99.38%
60	EU916482	99.69%
61	EU916479	99.07%
62	EU916478	99.69%
63	EU916477	99.07%
64	EU916476	99.69%
65	EU916474	99.38%
66	EU916473	99.69%
67	EU916472	99.38%
68	EU916471	99.38%
69	EU916470	99.69%
70	EU916468	99.69%
71	EU916465	99.69%
72	EU916381	99.07%
73	EU916380	99.69%
74	EU916377	99.38%
75	EU916376	99.69%
76	EU916375	99.07%
77	EU916374	99.07%
78	EU916373	99.69%
79	EU916370	99.38%
80	EU916369	99.69%
81	EU916368	100.00%
82	EU916367	99.69%
83	EU916366	99.38%
84	EU916365	99.69%
85	EU916364	99.07%
86	EU916363	99.69%
87	EU916362	99.07%
88	EU916361	99.38%
89	EU916358	99.69%

90	EU916357		99.38%
91	EU916356		99.38%
92	EU916355		99.38%
93	EU916352		99.69%
94	EU916348		99.38%
95	EU916347		99.69%
96	EU916346		98.77%
97	EU916345		99.69%
98	EU916344		99.07%
99	EU916343		99.38%
100	EU916342		99.38%
101	EU916341		99.07%
102	EU916340		99.69%
103	EU916339		99.38%
104	EU916338		99.69%
105	EU916337		99.38%
106	EU916336		99.69%
107	EU916335		99.38%
108	EU916334		99.38%
Cluster 487			
1	Contaminant	*	
Cluster 488			
1	HM801596	*	
2	HM801590		99.69%
3	HM801262		99.38%
4	HM801248		99.38%
5	HM801171		99.38%
6	HM801163		99.38%
7	HM801158		99.07%
8	HM801156		99.38%
9	HM801155		99.38%
Cluster 489			
1	HM801514	*	
2	HM801499		99.69%
3	HM801497		99.38%
4	HM801463		98.77%
5	HM801435		99.07%
6	HM801432		97.53%
7	HM801429		99.38%
8	HM801427		99.38%
9	HM801403		99.38%
10	HM801395		99.38%
11	HM801387		99.07%
12	HM801235		96.91%
Cluster 490			
1	HM801442	*	
2	HM801433		99.69%
Cluster 491			
1	HM801398	*	
2	HM801174		99.38%
Cluster 492			
1	HM801298	*	
Cluster 493			
1	HM801173	*	
2	HM801172		99.38%
3	HM801169		99.69%
4	HM801161		99.07%
5	HM801157		99.38%
6	6666666.181		97.84%
Cluster 494			
1	HM801170	*	
2	HM801165		99.69%
3	HM801164		98.77%
4	HM801152		99.38%
5	HM801151		99.69%
6	6666666.181		98.77%
Cluster 495			
1	KF151476	*	
Cluster 496			
1	KF151475	*	

Cluster 497				
	1	KF151474	*	
	2	KF151473		99.07%
	3	KF151471		99.38%
	4	KF151470		99.38%
	5	KF151467		99.38%
	6	KF151465		99.07%
	7	KF151464		99.07%
Cluster 498				
	1	KF151456	*	
	2	KF151455		99.69%
	3	KF151454		99.69%
Cluster 499				
	1	KF151448	*	
Cluster 500				
	1	KF151445	*	
Cluster 501				
	1	KF151427	*	
Cluster 502				
	1	KF151426	*	
	2	KF151425		99.38%
Cluster 503				
	1	KF151424	*	
Cluster 504				
	1	1085.;size=5;29	*	
Cluster 505				
	1	10853361;size=5	*	
Cluster 506				
	1	1121030.4	*	
Cluster 507				
	1	1134912.5	*	
Cluster 508				
	1	1206458.9	*	
Cluster 509				
	1	1228987.4	*	
Cluster 510				
	1	1261621.6	*	
Cluster 511				
	1	1337936.6	*	
Cluster 512				
	1	1337936.6	*	
Cluster 513				
	1	1538295.4	*	
Cluster 514				
	1	187420	*	
Cluster 515				
	1	188937.1	*	
Cluster 516				
	1	188937.1	*	
Cluster 517				
	1	188937.1	*	
Cluster 518				
	1	188937.1	*	
Cluster 519				
	1	216596	*	
Cluster 520				
	1	224911	*	
Cluster 521				
	1	240292.3	*	
Cluster 522				
	1	240292.3	*	
Cluster 523				
	1	240292.3	*	
Cluster 524				
	1	240292.3	*	
Cluster 525				
	1	243159.3	*	
	2	Acidithiobacillus		100.00%
Cluster 526				
	1	243233.4	*	

Cluster 527				
	1	264203.3	*	
Cluster 528				
	1	266834	*	
Cluster 529				
	1	266835	*	
Cluster 530				
	1	267377	*	
Cluster 531				
	1	272943.3	*	
	2	Rhodobacter		99.38%
Cluster 532				
	1	321327.2	*	
	2	321332		99.07%
Cluster 533				
	1	351627.4	*	
Cluster 534				
	1	354.43	*	
Cluster 535				
	1	357808.3	*	
Cluster 536				
	1	365044.32	*	
Cluster 537				
	1	36873	*	
Cluster 538				
	1	383372.4	*	
Cluster 539				
	1	391612.3	*	
	2	43989.3		98.46%
Cluster 540				
	1	395495.3	*	
Cluster 541				
	1	395961.4	*	
Cluster 542				
	1	395965.4	*	
Cluster 543				
	1	41431.3	*	
Cluster 544				
	1	414684.4	*	
Cluster 545				
	1	419665.8	*	
Cluster 546				
	1	431943.4	*	
Cluster 547				
	1	438753.3	*	
	2	438753.3		99.69%
Cluster 548				
	1	477974.3	*	
Cluster 549				
	1	481448.4	*	
Cluster 550				
	1	56107.12	*	
Cluster 551				
	1	56107.12	*	
Cluster 552				
	1	56107.12	*	
Cluster 553				
	1	579138.6	*	
Cluster 554				
	1	63737.4	*	
Cluster 555				
	1	63737.4	*	
Cluster 556				
	1	63737.4	*	
Cluster 557				
	1	65393.5	*	
Cluster 558				
	1	6666666.181	*	
	2	HQ611614		100.00%
	3	HQ611611		100.00%

	4	HQ611609		100.00%
	5	HQ611605		99.38%
	6	HQ611597		99.69%
	7	HQ611595		100.00%
Cluster 559				
	1	6666666.181	*	
Cluster 560				
	1	6666666.181	*	
Cluster 561				
	1	6666666.181	*	
Cluster 562				
	1	6666666.181	*	
Cluster 563				
	1	6666666.181	*	
Cluster 564				
	1	1609966.6	*	
Cluster 565				
	1	Methanosarcina	*	
Cluster 566				
	1	Methanosarcina	*	
Cluster 567				
	1	Methanococcus	*	
Cluster 568				
	1	Rhodopseudomonas	*	
	2	Rhodopseudomonas		97.22%
	3	EU916304		97.53%
	4	EU916283		97.22%
	5	KC140390		97.53%
	6	KC140389		96.91%
	7	KC140388		97.53%
	8	KC140386		96.91%
	9	KC140421		96.91%
	10	KC140420		96.91%
	11	KC140419		97.53%
	12	KC140414		96.60%
	13	KC140405		96.60%
	14	KC140404		97.22%
	15	KC140403		96.91%
	16	KC140402		96.91%
	17	KC140401		96.91%
	18	KC140400		97.22%
	19	KC140399		97.53%
	20	KC140396		96.60%
	21	KC140390		97.53%
	22	KC140389		96.91%
	23	KC140388		97.53%
	24	KC140386		96.91%
	25	KC140421		96.91%
	26	KC140420		96.91%
	27	KC140419		97.53%
	28	KC140414		96.60%
	29	KC140405		96.60%
	30	KC140404		97.22%
	31	KC140403		96.91%
	32	KC140402		96.91%
	33	KC140401		96.91%
	34	KC140400		97.22%
	35	KC140399		97.53%
	36	KC140396		96.60%
Cluster 569				
	1	Cylindrospermopsis	*	
Cluster 570				
	1	Rhodopseudomonas	*	
Cluster 571				
	1	Nostoc sp.	*	
Cluster 572				
	1	Rhodobacter	*	
Cluster 573				
	1	Burkholderia	*	
Cluster 574				

	1	HQ611951	*	
Cluster 575				
	1	HQ611947	*	
	2	HQ611945		99.69%
	3	HQ611937		100.00%
	4	HQ611936		99.07%
	5	HQ611935		98.50%
	6	HQ611927		100.00%
	7	HQ611915		97.84%
	8	HQ611907		99.07%
	9	HQ611904		97.84%
	10	HQ611903		99.69%
	11	HQ611902		97.53%
	12	HQ611899		100.00%
	13	HQ611898		97.84%
	14	HQ611897		98.46%
	15	HQ611896		97.84%
	16	HQ611895		98.46%
	17	HQ611894		100.00%
	18	HQ611891		100.00%
	19	HQ611856		98.46%
	20	HQ611830		99.69%
	21	HQ611806		98.50%
	22	HQ611795		99.38%
	23	HQ611793		100.00%
	24	HQ611787		100.00%
	25	HQ611758		100.00%
	26	HQ611687		98.50%
	27	HQ611659		97.22%
	28	HQ611535		98.50%
	29	HQ611484		97.84%
	30	HQ611459		99.69%
	31	HQ611453		98.46%
Cluster 576				
	1	HQ611933	*	
	2	HQ611890		99.38%
	3	HQ611838		98.46%
	4	AY896423		96.30%
Cluster 577				
	1	HQ611930	*	
Cluster 578				
	1	HQ611720	*	
	2	HQ611446		97.84%
Cluster 579				
	1	HQ611599	*	
	2	HQ611921		100.00%
	3	HQ611612		99.69%
	4	HQ611606		99.69%
	5	HQ611598		100.00%
	6	HQ611596		100.00%
Cluster 580				
	1	HQ611560	*	
Cluster 581				
	1	HQ611434	*	
Cluster 582				
	1	AB727501	*	
Cluster 583				
	1	AB727490	*	
Cluster 584				
	1	AB727488	*	
Cluster 585				
	1	HQ456024	*	
Cluster 586				
	1	HQ455983	*	
Cluster 587				
	1	HQ455952	*	
	2	HQ455948		99.69%
	3	HQ455939		99.07%
	4	HQ455937		99.69%
	5	HQ455932		99.69%

	6	HQ455925	99.38%
	7	HQ455922	99.69%
Cluster 588			
	1	HQ455911	*
Cluster 589			
	1	HQ455897	*
Cluster 590			
	1	EU916657	*
	2	EU916656	99.07%
	3	EU916652	98.77%
	4	EU916651	99.07%
	5	EU916649	99.07%
	6	EU916646	99.07%
	7	EU916644	99.38%
	8	EU916640	98.77%
	9	EU916632	98.77%
	10	EU916630	98.77%
	11	EU916623	99.38%
	12	EU916617	99.07%
	13	EU916559	98.77%
	14	EU916557	99.38%
	15	EU916552	98.77%
	16	EU916550	99.07%
	17	EU916547	99.38%
	18	EU916535	99.38%
	19	EU916527	99.38%
Cluster 591			
	1	EU916643	*
	2	EU916634	100.00%
	3	EU916616	100.00%
	4	EU916612	100.00%
Cluster 592			
	1	EU916629	*
	2	EU916619	99.69%
	3	EU916295	99.69%
	4	EU916292	99.69%
	5	KC140384	99.69%
	6	KC140382	99.69%
	7	KC140381	99.07%
	8	KC140380	99.38%
	9	KC140379	99.38%
	10	KC140378	99.69%
	11	KC140375	97.53%
	12	KC140413	99.38%
	13	KC140412	99.69%
	14	KC140409	99.38%
	15	KC140408	99.38%
	16	KC140407	99.69%
	17	KC140406	97.22%
	18	KC140365	99.07%
	19	KC140364	99.69%
	20	KC140361	99.69%
	21	KC140356	99.69%
	22	KC140384	99.69%
	23	KC140382	99.69%
	24	KC140381	99.07%
	25	KC140380	99.38%
	26	KC140379	99.38%
	27	KC140378	99.69%
	28	KC140375	97.53%
	29	KC140413	99.38%
	30	KC140412	99.69%
	31	KC140409	99.38%
	32	KC140408	99.38%
	33	KC140407	99.69%
	34	KC140406	97.22%
	35	KC140365	99.07%
	36	KC140364	99.69%
	37	KC140361	99.69%
	38	KC140356	99.69%

Cluster 593			
	1	EU916608	*
	2	EU916605	99.38%
	3	EU916567	99.69%
	4	EU916506	99.07%
	5	EU916481	99.38%
Cluster 594			
	1	EU916601	*
	2	EU916568	99.07%
	3	EU916435	97.84%
	4	EU916433	98.50%
	5	EU916424	97.84%
Cluster 595			
	1	EU916523	*
	2	EU916504	100.00%
	3	EU916499	100.00%
	4	EU916497	100.00%
Cluster 596			
	1	EU916490	*
Cluster 597			
	1	EU916489	*
Cluster 598			
	1	EU916462	*
	2	EU916457	99.69%
	3	EU916453	99.38%
	4	EU916449	99.38%
	5	EU916448	99.69%
	6	EU916443	99.38%
	7	EU916442	99.69%
	8	EU916439	99.38%
	9	EU916436	99.38%
	10	EU916434	99.07%
	11	EU916274	98.77%
Cluster 599			
	1	EU916460	*
	2	EU916426	100.00%
Cluster 600			
	1	EU916458	*
	2	EU916446	99.38%
	3	EU916444	99.69%
	4	EU916421	98.77%
	5	EU916332	98.46%
	6	EU916328	98.50%
Cluster 601			
	1	EU916455	*
	2	EU916447	99.69%
	3	EU916429	100.00%
	4	EU916378	100.00%
Cluster 602			
	1	EU916452	*
	2	KC140469	99.07%
	3	KC140460	98.46%
	4	KC140452	98.46%
	5	KC140449	98.77%
	6	KC140448	99.07%
	7	KC140445	98.77%
	8	KC140440	98.77%
	9	KC140426	99.07%
	10	KC140469	99.07%
	11	KC140460	98.46%
	12	KC140452	98.46%
	13	KC140449	98.77%
	14	KC140448	99.07%
	15	KC140445	98.77%
	16	KC140440	98.77%
	17	KC140426	99.07%
Cluster 603			
	1	EU916451	*
Cluster 604			
	1	EU916445	*

Cluster 605				
	1	EU916430	*	
	2	EU916427		99.07%
Cluster 606				
	1	EU916425	*	
Cluster 607				
	1	EU916416	*	
	2	EU916303		99.69%
Cluster 608				
	1	EU916415	*	
Cluster 609				
	1	EU916400	*	
	2	EU916399		99.07%
	3	EU916386		99.07%
Cluster 610				
	1	KC140385	*	
Cluster 611				
	1	KC140376	*	
Cluster 612				
	1	KC140415	*	
Cluster 613				
	1	KC140446	*	
	2	KC140442		99.69%
	3	KC140434		100.00%
	4	KC140425		100.00%
Cluster 614				
	1	AY972875	*	
Cluster 615				
	1	AY896365	*	
Cluster 616				
	1	HM801496	*	
	2	HM801507		99.38%
	3	HM801506		99.07%
	4	HM801490		99.07%
	5	HM801479		99.07%
Cluster 617				
	1	JQ358665	*	
	2	JQ358660		99.69%
Cluster 618				
	1	HM210357	*	
	2	HM210356		98.43%
	3	HM210355		97.51%
	4	HM210354		97.51%
	5	HM210353		97.51%
	6	HM210352		97.82%
	7	HM210350		97.51%
	8	HM210349		97.82%
	9	HM210348		97.82%
	10	HM210347		98.30%
	11	HM210346		97.51%
Cluster 619				
	1	HM210351	*	
Cluster 620				
	1	EF568436	*	
Cluster 621				
	1	EF568430	*	
Cluster 622				
	1	Desulfotomaculum nitrificans	*	
Cluster 623				
	1	10853579	*	
Cluster 624				
	1	203119	*	
Cluster 625				
	1	272562	*	
Cluster 626				
	1	272564b4	*	
Cluster 627				
	1	290402a34	*	
Cluster 628				
	1	290402b34	*	

Cluster 629				
	1	293826.4	*	
Cluster 630				
	1	349161.4	*	
Cluster 631				
	1	351160.3	*	
Cluster 632				
	1	431943.4	*	
	2	431943.4		98.44%
Cluster 633				
	1	431943.4	*	
Cluster 634				
	1	456442.1	*	
Cluster 635				
	1	456442.1	*	
Cluster 636				
	1	485916.4	*	
Cluster 637				
	1	485916.7	*	
Cluster 638				
	1	521011.3	*	
Cluster 639				
	1	521011.3	*	
Cluster 640				
	1	1123511.5	*	
Cluster 641				
	1	5803274	*	
Cluster 642				
	1	HQ456023	*	
Cluster 643				
	1	HQ455860	*	
Cluster 644				
	1	HQ455844	*	
Cluster 645				
	1	EU916417	*	
	2	EU916413		99.69%
	3	EU916408		99.38%
	4	EU916398		99.69%
	5	EU916396		99.07%
	6	EU916395		99.69%
	7	EU916388		99.38%
	8	EU916385		99.69%
	9	EU916383		99.69%
Cluster 646				
	1	EU916333	*	
	2	EU916331		99.69%
Cluster 647				
	1	HM210368	*	
Cluster 648				
	1	EU052544	*	
Cluster 649				
	1	EF568543	*	
Cluster 650				
	1	323259.5	*	
Cluster 651				
	1	368407.6	*	
Cluster 652				
	1	521011b3	*	
Cluster 653				
	1	9903167	*	
Cluster 654				
	1	AY896317	*	

Supplemental Table 10: Environmental and hydrographic parameters and isolate *nifH* abundances throughout the AZMP, Scotian Shelf and GEOVIDE sampling.

Cruise	Statio ID	Latitude (°N)	Longitude (°E)	Depth (m)	Salinity (PSU)	Temp (°C)	O_2 (µM)	Chloro (mg/m^3)	$NO3^-$ (mM)	PO_4^{3-} (mM)	SiO_4^{4-} (mM)	NH3 (mM)	NO_2^- (mM)	N* (mM)	isolate (*nifH* copies L^{-1})
AZMP	BBL1_1	43.250	-65.480	1	31.30	1.76	344.02	1.718	5.11	0.70	5.65	0.84	0.12	-3.14	68
AZMP	BBL1_10	43.250	-65.480	10	31.30	1.75	344.54	1.902	5.12	0.70	5.42	0.78	0.12	-3.08	152
AZMP	BBL1_20	43.250	-65.480	20	31.30	1.76	343.82	1.887	5.30	0.70	5.42	0.78	0.12	-2.94	90
AZMP	BBL1_40	43.250	-65.480	40	31.61	1.95	333.52	1.687	5.73	0.72	5.41	0.78	0.11	-2.84	64
AZMP	BBL2_1	43.000	-65.481	1	31.46	1.90	349.03	3.651	3.78	0.61	3.38	0.74	0.11	-3.03	331
AZMP	BBL2_20	43.000	-65.481	20	31.46	1.91	348.55	3.353	3.70	0.60	4.09	0.78	0.11	-2.82	174
AZMP	BBL2_40	43.000	-65.481	40	31.69	1.94	343.67	2.929	3.46	0.57	2.61	1.21	0.12	-2.70	121
AZMP	BBL2_80	43.000	-65.481	80	32.34	3.51	288.03	0.782	7.83	0.86	7.10	1.01	0.14	-2.87	33
AZMP	BBL5_1	42.133	-65.501	1	33.76	7.80	291.35	2.886	4.16	0.48	1.52	1.06	0.17	-0.48	97
AZMP	BBL5_20	42.133	-65.501	20	34.05	8.44	275.18	2.378	6.38	0.57	2.42	1.18	0.29	0.41	41
AZMP	BBL5_40	42.133	-65.501	40	34.17	8.72	267.59	2.071	7.21	0.61	2.80	1.59	0.22	0.56	30
AZMP	BBL5_80	42.133	-65.501	80	35.13	11.20	219.80	0.638	11.84	0.76	5.78	1.01	0.39	2.90	80
AZMP	BBL7_1	41.866	-65.349	1	33.43	6.91	300.38	3.396	5.52	0.59	2.45	0.93	0.14	-0.83	97
AZMP	BBL7_20	41.866	-65.349	20	33.67	7.03	270.90	2.243	9.37	0.81	4.65	0.71	0.13	-0.50	92
AZMP	BBL7_250	41.866	-65.349	250	35.30	10.77	165.61	0.000	19.19	1.24	10.01	0.46	0.04	2.37	47
AZMP	BBL7_80	41.866	-65.349	80	35.27	11.98	243.21	0.598	8.36	0.60	3.08	0.68	0.36	2.04	477
AZMP	CSL1_1	46.958	-60.216	1	29.85	-0.13	402.09	9.352	-0.07	0.38	0.12	0.71	0.07	-3.15	27
AZMP	CSL1_20	46.958	-60.216	20	30.03	-0.37	399.44	10.564	0.30	0.38	0.16	0.55	0.00	-2.93	22
AZMP	CSL1_40	46.958	-60.216	40	31.04	-1.09	387.04	14.201	1.69	0.57	1.94	0.68	0.06	-4.46	5
AZMP	CSL1_60	46.958	-60.216	60	31.51	-0.57	352.61	8.313	4.37	0.72	4.32	1.21	0.09	-4.10	12
AZMP	CSL4_1	47.270	-59.784	1	30.47	-0.43	410.87	12.469	0.00	0.46	0.15	0.43	0.00	-4.38	27
AZMP	CSL4_20	47.270	-59.784	20	30.77	-1.03	391.81	15.760	0.73	0.49	0.94	0.80	0.03	-4.18	27
AZMP	CSL4_300	47.270	-59.784	300	34.74	5.99	139.14	0.000	23.21	1.66	25.67	0.21	0.00	-0.42	8
AZMP	CSL4_60	47.270	-59.784	60	31.64	-1.47	351.97	0.813	6.36	0.83	6.70	0.80	0.13	-3.96	5
AZMP	CSL6_1	47.579	-59.342	1	31.26	-0.60	389.63	8.833	1.83	0.57	3.24	0.69	0.09	-4.31	2457
AZMP	CSL6_20	47.579	-59.342	20	31.77	-0.27	363.43	2.209	4.91	0.72	5.16	0.94	0.13	-3.56	462
AZMP	CSL6_200	47.579	-59.342	200	34.54	7.05	188.16	0.000	17.23	1.26	13.75	0.56	0.00	-0.06	1043
AZMP	CSL6_60	47.579	-59.342	60	32.30	0.34	343.66	0.260	6.28	0.77	5.96	0.88	0.16	-2.95	299
AZMP	GULD04_1	43.790	-58.901	1	32.28	1.77	358.81	3.184	2.34	0.51	1.97	0.92	0.16	-2.76	54
AZMP	GULD04_100	43.790	-58.901	100	32.58	2.69	341.32	0.936	2.43	0.47	1.79	1.25	0.15	-2.06	52
AZMP	GULD04_20	43.790	-58.901	20	34.81	7.49	190.95	2.178	18.40	1.23	10.27	0.72	0.03	1.67	223
AZMP	GULD04_250	43.790	-58.901	250	35.05	11.49	229.87	0.000	8.68	0.63	3.90	1.02	0.25	1.78	33
AZMP	HL1_1	44.400	-63.450	1	31.29	0.30	394.07	11.084	-0.03	0.41	0.00	0.52	0.03	-3.63	51
AZMP	HL1_20	44.400	-63.450	20	31.29	0.24	393.63	11.777	-0.04	0.42	0.00	0.58	0.04	-3.82	69
AZMP	HL1_40	44.400	-63.450	40	31.62	0.57	358.62	5.730	2.86	0.62	1.78	1.24	0.09	-4.11	90

AZMP	HL1_60	44.400	-63.450	60	32.75	4.12	298.24	0.706	5.09	0.69	3.60	1.68	0.10	-2.95	66
AZMP	HL2_1	44.267	-63.317	1	31.23	-0.58	398.08	9.698	0.64	0.46	0.60	0.68	0.05	-3.77	33
AZMP	HL2_20	44.267	-63.317	20	31.27	-0.80	391.90	10.564	0.96	0.50	0.92	0.58	0.07	-4.05	36
AZMP	HL2_40	44.267	-63.317	40	31.29	-0.82	376.59	8.192	2.14	0.57	1.50	0.64	0.08	-3.98	5
AZMP	HL2_80	44.267	-63.317	80	31.40	-0.40	371.93	7.683	2.51	0.59	1.78	1.40	0.09	-3.91	22
AZMP	HL4_1	43.480	-62.452	1	32.77	5.36	311.51	0.319	2.04	0.48	0.38	2.00	0.09	-2.67	56
AZMP	HL4_20	43.480	-62.452	20	32.79	5.36	311.34	0.356	2.28	0.50	0.68	2.17	0.10	-2.67	29
AZMP	HL4_40	43.480	-62.452	40	32.98	5.61	298.65	0.209	3.05	0.53	0.86	2.63	0.13	-2.35	61
AZMP	HL4_60	43.480	-62.452	60	34.31	9.26	202.44	0.128	12.06	1.08	6.56	2.74	0.19	-2.09	24
AZMP	HL5.5_1	42.940	-61.831	1	33.29	6.94	309.01	2.025	2.78	0.46	1.21	0.97	0.16	-1.49	85
AZMP	HL5.5_20	42.940	-61.831	20	33.38	7.02	305.03	1.519	2.44	0.45	0.96	0.95	0.16	-1.70	28
AZMP	HL5.5_250	42.940	-61.831	250	35.29	10.14	137.45	0.000	23.44	1.51	12.21	0.67	0.03	2.24	32
AZMP	HL5.5_80	42.940	-61.831	80	35.03	11.22	238.48	0.524	9.56	0.66	3.10	1.37	0.33	2.24	14
AZMP	HL8_1	42.363	-61.345	1	33.53	7.91	320.38	2.454	0.08	0.26	0.00	0.69	0.02	-1.21	17
AZMP	HL8_100	42.363	-61.345	100	35.51	13.01	220.60	0.260	9.92	0.64	3.96	0.00	0.00	2.56	56
AZMP	HL8_20	42.363	-61.345	20	33.53	7.86	318.79	2.439	0.18	0.25	0.00	0.00	0.00	-0.90	12
AZMP	HL8_250	42.363	-61.345	250	35.27	10.17	144.41	0.000	18.76	1.29	9.71	0.00	0.00	0.97	8
AZMP	LHB2_1	44.086	-63.903	1	31.23	1.20	382.93	8.659	0.30	0.41	0.25	0.69	0.04	-3.26	207
AZMP	LHB2_20	44.086	-63.903	20	31.40	0.58	378.57	10.564	1.02	0.48	0.08	1.05	0.06	-3.73	119
AZMP	LHB2_40	44.086	-63.903	40	31.44	0.30	370.51	10.911	1.70	0.55	0.56	1.58	0.10	-4.16	37
AZMP	LHB2_80	44.086	-63.903	80	31.91	1.33	338.74	2.301	3.99	0.69	2.98	1.67	0.09	-4.00	42
AZMP	LHB4_1	43.379	-63.667	1	32.15	3.13	340.72	1.841	1.65	0.44	0.76	1.57	0.07	-2.44	102
AZMP	LHB4_20	43.379	-63.667	20	32.24	3.34	336.17	1.657	1.76	0.44	0.84	1.18	0.07	-2.37	219
AZMP	LHB4_40	43.379	-63.667	40	32.27	3.30	330.52	1.626	2.07	0.48	1.07	1.37	0.07	-2.58	98
AZMP	LHB4_80	43.379	-63.667	80	33.62	6.67	203.33	0.997	13.65	1.16	10.30	1.80	0.15	-1.89	80
AZMP	LHB6.0_1	42.666	-63.415	1	34.62	10.43	286.52	2.802	3.49	0.36	0.58	0.65	0.23	0.81	44
AZMP	LHB6.0_20	42.666	-63.415	20	34.64	10.47	285.43	2.730	3.61	0.37	0.68	1.10	0.27	0.82	34
AZMP	LHB6.0_250	42.666	-63.415	250	35.40	11.06	138.27	0.000	22.41	1.40	12.12	0.48	0.04	2.93	875
AZMP	LHB6.0_80	42.666	-63.415	80	35.24	11.92	241.57	0.656	8.18	0.59	2.78	0.95	0.41	1.99	38
AZMP	LHB6.7_1	42.193	-63.252	1	35.74	13.38	239.93	0.492	9.81	0.58	3.68	0.63	0.17	3.60	118
AZMP	LHB6.7_20	42.193	-63.252	20	35.76	13.37	238.86	0.460	9.94	0.58	3.68	0.53	0.17	3.70	38
AZMP	LHB6.7_250	42.193	-63.252	250	35.86	13.63	228.96	0.000	8.88	0.60	3.48	0.62	0.13	2.38	45
AZMP	LHB6.7_80	42.193	-63.252	80	35.78	13.35	236.64	0.433	9.29	0.58	3.62	0.59	0.17	3.13	127
AZMP	LL4_1	45.158	-59.175	1	31.69	1.16	366.22	4.669	3.65	0.67	2.34	1.26	0.15	-4.02	96
AZMP	LL4_20	45.158	-59.175	20	31.69	1.14	366.33	5.051	3.63	0.66	2.36	0.86	0.14	-3.96	97
AZMP	LL4_40	45.158	-59.175	40	31.69	1.12	363.52	4.882	3.20	0.62	1.94	1.35	0.12	-3.76	27
AZMP	LL4_80	45.158	-59.175	80	32.13	1.25	310.57	1.166	8.38	0.98	9.20	0.85	0.15	-4.29	376
AZMP	LL7_1	44.132	-58.175	1	32.45	3.34	358.62	5.391	0.33	0.34	0.48	0.56	0.04	-2.21	191
AZMP	LL7_20	44.132	-58.175	20	32.45	3.33	357.42	5.688	-0.04	0.28	0.54	0.49	0.04	-1.53	244
AZMP	LL7_250	44.132	-58.175	250	34.92	7.50	180.20	0.000	20.98	1.28	12.93	1.44	0.05	3.41	26
AZMP	LL7_80	44.132	-58.175	80	33.01	4.53	313.02	0.644	3.01	0.44	2.27	1.51	0.14	-0.95	156
AZMP	LL9_1	43.469	-57.531	1	32.46	3.79	365.53	2.420	0.42	0.32	0.42	0.58	0.00	-1.83	16
AZMP	LL9_20	43.469	-57.531	20	32.47	3.81	360.37	2.674	0.30	0.32	0.41	0.64	0.00	-1.89	12
AZMP	LL9_250	43.469	-57.531	250	34.87	6.67	192.96	0.000	18.89	1.28	12.04	1.14	0.06	1.41	3

AZMP	LL9_80	43.469	-57.531	80	33.00	4.53	315.70	0.442	2.38	0.47	1.33	1.26	0.09	-2.15	16
AZMP	STAB01_1	45.999	-59.529	1	30.61	-0.12	411.11	14.028	0.05	0.40	0.00	0.85	0.06	-3.37	423
AZMP	STAB01_10	45.999	-59.529	10	30.65	-0.21	412.61	14.894	0.07	0.42	0.00	0.84	0.05	-3.64	247
AZMP	STAB01_20	45.999	-59.529	20	30.78	-0.39	415.01	17.492	0.20	0.43	0.00	0.70	0.00	-3.80	142
AZMP	STAB01_40	45.999	-59.529	40	31.15	-0.94	370.90	6.537	3.44	0.69	3.58	1.20	0.09	-4.68	197
AZMP	STAB05_1	46.420	-58.875	1	31.18	-0.51	406.29	13.509	0.52	0.47	1.00	0.57	0.00	-4.15	348
AZMP	STAB05_20	46.420	-58.875	20	31.18	-0.50	404.59	14.894	0.80	0.50	1.02	0.56	0.00	-4.24	126
AZMP	STAB05_300	46.420	-58.875	300	34.75	6.10	148.70	0.000	23.00	1.66	25.12	0.71	0.00	-0.69	241
AZMP	STAB05_80	46.420	-58.875	80	31.85	-0.63	361.13	0.178	5.51	0.76	5.35	0.84	0.15	-3.58	324
AZMP	BBL1_1	43.250	-65.481	1	31.07	15.27	255.02	0.550	1.46	0.19	1.00	0.40	0.00	1.26	76
AZMP	BBL1_10	43.250	-65.481	10	31.07	15.25	254.36	0.566	0.35	0.18	1.04	0.51	0.00	0.45	19
AZMP	BBL1_20	43.250	-65.481	20	31.13	14.53	254.37	0.401	0.41	0.26	1.30	0.69	0.06	-0.74	31
AZMP	BBL1_40	43.250	-65.481	40	31.88	9.55	269.63	0.271	2.25	0.52	3.15	0.81	0.15	-3.00	25
AZMP	BBL3_1	42.760	-65.483	1	33.00	16.80	249.17	0.491	-0.03	0.18	1.04	0.30	0.03	-0.01	83
AZMP	BBL3_20	42.760	-65.483	20	33.27	16.68	245.77	1.108	0.08	0.22	1.20	0.40	0.06	-0.43	21
AZMP	BBL3_40	42.760	-65.483	40	33.82	13.59	227.70	0.299	5.18	0.55	3.45	0.46	0.19	-0.47	38
AZMP	BBL3_80	42.760	-65.483	80	34.03	12.46	214.36	0.127	7.71	0.70	5.22	0.38	0.15	-0.39	26
AZMP	BBL5_1	42.132	-65.499	1	34.46	19.59	232.12	0.354	0.00	0.10	1.12	0.37	0.00	1.36	20
AZMP	BBL5_20	42.132	-65.499	20	34.45	19.58	231.37	0.358	0.00	0.10	1.14	0.35	0.00	1.27	28
AZMP	BBL5_40	42.132	-65.499	40	34.44	19.49	231.29	0.456	0.00	0.11	1.16	0.36	0.00	1.08	23
AZMP	BBL5_80	42.132	-65.499	80	35.33	14.71	183.34	0.115	8.31	0.61	4.22	0.51	0.00	1.50	20
AZMP	BBL7_1	41.867	-65.349	1	34.67	21.02	227.98	0.271	1.13	0.10	0.96	0.34	0.00	2.48	67
AZMP	BBL7_20	41.867	-65.349	20	34.67	21.03	227.61	0.271	0.00	0.10	1.18	0.37	0.00	1.33	58
AZMP	BBL7_250	41.867	-65.349	250	35.45	11.40	140.98	0.000	21.37	1.36	10.72	0.36	0.00	2.59	28
AZMP	BBL7_80	41.867	-65.349	80	34.66	15.61	218.94	0.173	4.32	0.41	2.96	0.32	0.15	0.82	23
AZMP	CSL1_1	46.960	-60.218	1	28.91	14.34	263.10	2.129	-0.02	0.20	0.56	0.63	0.02	-0.27	41
AZMP	CSL1_20	46.960	-60.218	20	29.04	14.02	261.02	1.254	0.05	0.25	0.74	1.54	0.05	-1.00	34
AZMP	CSL1_40	46.960	-60.218	40	31.13	3.37	281.09	0.149	6.04	1.02	10.90	3.51	0.30	-7.03	22
AZMP	CSL1_60	46.960	-60.218	60	31.97	1.16	287.70	0.099	8.09	1.13	11.72	2.97	0.22	-6.90	49
AZMP	CSL4_1	47.271	-59.782	1	30.98	9.38	301.73	1.663	0.00	0.31	0.80	1.00	0.00	-2.09	70
AZMP	CSL4_20	47.271	-59.782	20	31.36	5.89	325.13	1.429	1.37	0.55	1.92	0.74	0.12	-4.41	17
AZMP	CSL4_300	47.271	-59.782	300	34.79	6.01	143.45	0.000	22.05	1.65	21.74	0.27	0.00	-1.48	110
AZMP	CSL4_60	47.271	-59.782	60	32.34	0.97	314.87	0.033	7.27	0.91	5.36	0.42	0.12	-4.20	39
AZMP	CSL6_1	47.583	-59.344	1	31.09	9.24	296.28	0.963	0.20	0.33	1.13	0.36	0.00	-2.10	56
AZMP	CSL6_20	47.583	-59.344	20	31.37	7.17	310.04	0.977	0.71	0.40	1.43	0.39	0.00	-2.84	28
AZMP	CSL6_200	47.583	-59.344	200	34.05	5.50	200.18	0.000	15.66	1.28	15.14	0.32	0.00	-1.94	13
AZMP	CSL6_60	47.583	-59.344	60	32.26	1.37	325.55	0.111	4.51	0.76	3.64	0.98	0.14	-4.61	78
AZMP	GULD04_1	43.789	-58.900	1	31.34	16.59	255.15	0.216	0.00	0.14	0.36	0.17	0.00	0.74	66
AZMP	GULD04_100	43.789	-58.900	100	33.69	6.40	252.17	0.026	8.74	0.90	6.58	0.44	0.00	-2.81	55
AZMP	GULD04_20	43.789	-58.900	20	31.44	16.39	255.67	0.255	0.00	0.15	0.38	0.00	0.00	0.52	2866
AZMP	GULD04_250	43.789	-58.900	250	35.29	10.82	159.27	0.000	19.22	1.28	10.24	0.40	0.00	1.62	73
AZMP	HL1_1	44.400	-63.450	1	30.53	17.20	246.37	0.546	0.00	0.18	1.14	0.41	0.00	-0.01	148
AZMP	HL1_20	44.400	-63.450	20	30.54	17.22	246.07	0.499	0.00	0.17	0.70	0.36	0.00	0.15	264
AZMP	HL1_40	44.400	-63.450	40	31.48	6.94	285.31	0.334	2.56	0.64	3.64	1.49	0.20	-4.61	123

AZMP	HL1_60	44.400	-63.450	60	32.26	4.75	284.53	0.138	4.37	0.74	4.05	0.45	0.21	-4.39	233
AZMP	HL11_1	41.775	-60.905	1	36.02	23.67	212.46	0.156	0.00	0.01	0.85	0.73	0.00	2.68	270
AZMP	HL11_20	41.775	-60.905	20	36.03	23.69	213.50	0.159	0.00	0.03	0.72	0.27	0.00	2.50	114
AZMP	HL11_250	41.775	-60.905	250	35.82	14.07	166.94	0.000	13.96	0.87	5.65	0.36	0.02	3.01	184
AZMP	HL11_80	41.775	-60.905	80	36.38	19.23	186.26	0.177	4.19	0.26	2.13	0.35	0.08	3.07	44
AZMP	HL2_1	44.266	-63.319	1	30.57	17.61	243.92	0.234	0.00	0.10	0.46	0.43	0.00	1.30	102
AZMP	HL2_20	44.266	-63.319	20	31.22	16.98	251.31	0.409	0.00	0.19	0.46	0.46	0.00	-0.14	249
AZMP	HL2_40	44.266	-63.319	40	32.37	4.83	280.87	0.140	4.55	0.74	4.54	0.38	0.14	-4.20	76
AZMP	HL2_80	44.266	-63.319	80	32.83	4.40	250.73	0.042	8.30	0.98	8.86	0.42	0.00	-4.40	77
AZMP	HL4_1	43.481	-62.449	1	31.63	18.16	240.44	0.189	0.00	0.07	0.44	0.39	0.00	1.78	563
AZMP	HL4_20	43.481	-62.449	20	31.96	18.40	240.11	0.242	-0.02	0.07	0.55	0.32	0.02	1.75	261
AZMP	HL4_40	43.481	-62.449	40	32.85	7.71	300.57	1.021	-0.04	0.38	1.75	0.38	0.04	-3.10	103
AZMP	HL4_60	43.481	-62.449	60	33.18	6.16	259.56	0.119	5.08	0.75	4.24	0.33	0.06	-3.99	118
AZMP	HL6_1	42.834	-61.732	1	33.93	19.97	235.54	0.232	-0.02	0.03	0.70	0.28	0.02	2.40	261
AZMP	HL6_20	42.834	-61.732	20	34.11	20.16	234.37	0.228	-0.03	0.04	0.62	0.28	0.03	2.24	140
AZMP	HL6_250	42.834	-61.732	250	35.14	8.27	154.39	0.000	24.01	1.59	14.66	0.30	0.04	1.50	84
AZMP	HL6_80	42.834	-61.732	80	34.23	10.96	224.60	0.055	7.28	0.67	4.36	0.25	0.04	-0.55	103
AZMP	HL8_1	42.363	-61.341	1	35.03	20.97	231.76	0.224	-0.03	0.03	0.64	0.30	0.03	2.39	57
AZMP	HL8_100	42.363	-61.341	100	35.78	15.17	186.06	0.050	9.21	0.59	3.27	0.25	0.05	2.73	85
AZMP	HL8_20	42.363	-61.341	20	35.03	20.97	231.56	0.236	-0.02	0.04	0.60	0.30	0.02	2.24	93
AZMP	HL8_250	42.363	-61.341	250	35.52	12.03	152.00	0.000	19.45	1.23	9.14	0.26	0.05	2.76	58
AZMP	LL4_1	45.152	-59.173	1	30.18	15.89	253.74	0.389	0.00	0.14	0.20	0.26	0.00	0.66	19
AZMP	LL4_20	45.152	-59.173	20	30.28	15.75	253.79	0.471	0.00	0.15	0.19	0.34	0.00	0.44	46
AZMP	LL4_40	45.152	-59.173	40	32.02	3.04	317.07	0.193	3.64	0.72	2.96	1.30	0.18	-4.72	39
AZMP	LL4_80	45.152	-59.173	80	32.48	2.22	282.94	0.036	8.27	0.95	8.55	0.29	0.00	-4.00	50
AZMP	LL7_1	44.130	-58.182	1	32.15	17.03	248.77	0.373	0.26	0.11	0.35	0.33	0.00	1.34	31
AZMP	LL7_20	44.130	-58.182	20	32.24	17.13	248.96	0.385	0.00	0.12	0.35	0.55	0.00	0.96	110
AZMP	LL7_250	44.130	-58.182	250	35.26	9.93	139.39	0.000	22.92	1.46	12.63	0.29	0.00	2.40	12
AZMP	LL7_80	44.130	-58.182	80	33.85	8.73	251.29	0.151	5.02	0.61	3.69	0.46	0.12	-1.77	26
AZMP	LL9_1	43.474	-57.526	1	35.51	20.93	226.02	0.240	-0.07	0.01	0.49	0.34	0.07	2.74	65
AZMP	LL9_20	43.474	-57.526	20	35.51	20.94	225.69	0.240	0.00	0.01	0.51	0.30	0.00	2.68	45
AZMP	LL9_250	43.474	-57.526	250	35.65	13.34	186.22	0.000	12.97	0.81	5.26	0.30	0.03	3.02	46
AZMP	LL9_80	43.474	-57.526	80	36.02	16.97	194.80	0.153	4.13	0.29	1.77	0.27	0.13	2.46	21
AZMP	STAB01_1	45.996	-59.533	1	29.12	14.13	267.02	1.677	-0.06	0.20	0.64	0.39	0.06	-0.22	20
AZMP	STAB01_10	45.996	-59.533	10	29.13	14.11	266.32	1.677	-0.07	0.20	0.68	0.52	0.07	-0.32	79
AZMP	STAB01_20	45.996	-59.533	20	29.35	13.61	262.08	0.963	-0.08	0.24	0.99	0.84	0.08	-0.97	86
AZMP	STAB01_40	45.996	-59.533	40	30.17	11.46	272.85	0.496	0.91	0.41	2.19	1.45	0.14	-2.62	54
AZMP	STAB05_1	46.416	-58.884	1	30.41	13.28	271.24	0.685	0.00	0.18	0.26	0.32	0.00	0.05	43
AZMP	STAB05_20	46.416	-58.884	20	30.81	10.68	278.39	0.861	0.22	0.26	0.62	0.43	0.05	-1.06	31
AZMP	STAB05_300	46.416	-58.884	300	34.83	5.76	133.73	0.000	22.40	1.81	32.15	0.51	0.04	-3.57	64
AZMP	STAB05_80	46.416	-58.884	80	32.68	1.59	291.11	0.036	9.59	1.00	7.90	0.68	0.00	-3.43	53
Bedford B.	2014-01-15	44.690	-63.640	1	27.02	3.60	315.31	0.610	9.00	0.82	15.92	5.46	0.37	-1.22	51
Bedford B.	2014-01-15	44.690	-63.640	5	28.87	3.45	304.08	0.490	8.13	0.91	11.56	4.09	0.26	-3.53	0
Bedford B.	2014-01-15	44.690	-63.640	10	29.52	3.34	289.70	0.550	7.82	0.97	10.17	3.21	0.23	-4.80	0

Bedford B.	2014-01-15	44.690	-63.640	60	31.23	5.25	69.78	0.060	16.65	3.34	34.31	0.55	0.07	-33.89	0
Bedford B.	2014-01-23	44.690	-63.640	1	27.97	1.04	329.23	0.640	8.39	1.00	11.72	4.01	0.25	-4.71	0
Bedford B.	2014-01-23	44.690	-63.640	5	29.54	2.49	290.46	0.670	8.20	1.15	10.96	2.92	0.25	-7.30	107
Bedford B.	2014-01-23	44.690	-63.640	10	30.19	3.05	282.32	0.460	8.15	1.11	10.24	1.90	0.24	-6.71	0
Bedford B.	2014-01-23	44.690	-63.640	60	31.23	5.24	61.70	0.230	17.13	3.98	37.78	1.04	0.33	-43.65	88
Bedford B.	2014-01-29	44.690	-63.640	1	28.76	1.13	525.47	0.720	8.60	1.07	12.07	3.22	0.24	-5.62	0
Bedford B.	2014-01-29	44.690	-63.640	5	29.53	2.22	505.14	0.810	8.68	1.07	11.99	3.22	0.22	-5.54	0
Bedford B.	2014-01-29	44.690	-63.640	10	29.74	2.45	479.07	0.610	8.30	1.07	11.22	2.51	0.21	-5.92	0
Bedford B.	2014-01-29	44.690	-63.640	60	31.03	4.63	146.53	0.090	14.49	3.21	29.93	0.54	0.16	-33.97	122
Bedford B.	2014-02-04	44.690	-63.640	1	28.30	1.41	738.69	1.270	8.97	1.03	11.60	5.55	0.24	-4.61	82
Bedford B.	2014-02-04	44.690	-63.640	5	29.59	2.32	-85.60	1.530	8.86	1.02	10.91	3.04	0.21	-4.56	0
Bedford B.	2014-02-04	44.690	-63.640	10	29.95	2.29	-85.50	0.520	8.52	1.03	10.20	2.72	0.17	-5.06	150
Bedford B.	2014-02-04	44.690	-63.640	60	30.88	4.58	274.41	0.090	16.89	3.01	29.45	0.71	0.14	-28.37	476
Bedford B.	2014-02-12	44.690	-63.640	1	29.20	0.29	334.44	1.010	8.86	1.09	11.01	5.16	0.24	-5.68	6400
Bedford B.	2014-02-12	44.690	-63.640	5	29.94	1.87	311.44	1.320	8.65	1.05	10.54	4.30	0.23	-5.25	9346
Bedford B.	2014-02-12	44.690	-63.640	10	30.18	2.34	299.31	0.950	8.60	1.05	10.32	2.10	0.20	-5.30	32783
Bedford B.	2014-02-12	44.690	-63.640	60	30.91	4.52	96.99	0.070	16.85	3.04	27.52	0.50	0.15	-28.89	6557
Bedford B.	2014-02-19	44.690	-63.640	1	28.91	0.98	328.96	1.030	9.09	0.85	9.81	2.65	0.19	-1.61	2877
Bedford B.	2014-02-19	44.690	-63.640	5	28.95	0.94	328.61	1.240	9.11	0.83	10.18	2.53	0.18	-1.27	10970
Bedford B.	2014-02-19	44.690	-63.640	10	29.32	1.26	316.91	1.270	9.18	0.89	9.90	2.54	0.19	-2.16	12152
Bedford B.	2014-02-19	44.690	-63.640	60	30.94	4.13	135.90	0.080	16.03	1.97	27.34	0.54	0.07	-12.59	6649
Bedford B.	2014-02-26	44.690	-63.640	1	28.91	0.96	554.61	2.120	9.29	0.80	9.22	2.75	0.22	-0.61	3841
Bedford B.	2014-02-26	44.690	-63.640	5	29.13	1.58	317.58	2.500	8.71	0.83	9.04	2.42	0.20	-1.67	4218
Bedford B.	2014-02-26	44.690	-63.640	10	30.09	1.47	307.55	1.400	8.69	0.96	9.39	2.10	0.19	-3.77	1341829
Bedford B.	2014-02-26	44.690	-63.640	60	30.94	4.00	136.24	0.070	16.30	2.13	28.89	0.42	0.17	-14.88	3763
Bedford B.	2014-03-05	44.690	-63.640	1	29.30	0.36	340.25	3.820	7.81	0.76	7.89	1.98	0.20	-1.45	2032
Bedford B.	2014-03-05	44.690	-63.640	5	29.45	0.67	331.95	3.820	7.81	0.76	8.30	1.92	0.20	-1.45	7665
Bedford B.	2014-03-05	44.690	-63.640	10	30.26	1.51	306.34	2.160	8.51	0.85	9.09	1.67	0.18	-2.19	4577
Bedford B.	2014-03-05	44.690	-63.640	60	30.90	3.43	160.60	0.110	15.26	2.11	27.62	0.80	0.16	-15.60	1854
Bedford B.	2014-03-12	44.690	-63.640	1	29.16	0.91	376.41	17.670	1.89	0.53	3.10	5.71	0.26	-3.69	1825
Bedford B.	2014-03-12	44.690	-63.640	5	30.21	1.48	322.68	10.220	4.44	0.64	5.23	1.80	0.24	-2.90	4044
Bedford B.	2014-03-12	44.690	-63.640	10	30.45	1.65	302.85	2.970	8.13	0.79	8.66	1.56	0.23	-1.61	11764
Bedford B.	2014-03-12	44.690	-63.640	60	30.76	1.54	282.43	0.830	9.15	1.12	11.97	1.60	0.17	-5.87	2486
Bedford B.	2014-03-19	44.690	-63.640	10	30.28	1.64	305.84	4.840	7.19	0.86	7.65	1.87	0.20	-3.67	18672
Bedford B.	2014-03-19	44.690	-63.640	60	30.81	1.35	298.85	1.700	7.82	0.95	9.75	1.79	0.17	-4.48	1153
Bedford B.	2014-03-28	44.690	-63.640	1	30.31	1.23	312.06	1.530	7.54	0.85	8.47	1.77	0.15	-3.16	85
Bedford B.	2014-03-28	44.690	-63.640	5	30.32	1.23	311.70	2.160	7.42	0.84	8.27	1.41	0.15	-3.12	1280
Bedford B.	2014-03-28	44.690	-63.640	10	30.38	1.30	307.15	2.500	6.78	0.93	7.80	1.39	0.13	-5.20	14321
Bedford B.	2014-03-28	44.690	-63.640	60	31.10	1.58	293.70	0.710	7.63	0.91	9.35	1.54	0.15	-4.03	5946
Bedford B.	2014-04-02	44.690	-63.640	1	28.91	1.26	330.11	0.750	7.74	0.78	8.68	2.21	0.18	-1.84	911

Bedford B.	2014-04-02	44.690	-63.640	5	29.34	1.32	320.89	0.720	7.68	0.76	8.49	2.32	0.17	-1.58	4858
Bedford B.	2014-04-02	44.690	-63.640	10	30.36	1.50	308.76	1.830	7.31	0.81	7.98	1.70	0.16	-2.75	1953
Bedford B.	2014-04-02	44.690	-63.640	60	31.07	1.58	286.62	0.260	7.70	0.95	9.21	1.85	0.15	-4.60	6023
Bedford B.	2014-04-09	44.690	-63.640	1	28.03	3.26	330.79	0.720	10.48	0.75	17.16	5.56	0.44	1.38	2522
Bedford B.	2014-04-09	44.690	-63.640	5	29.25	2.43	321.26	3.230	6.61	0.82	9.96	1.77	0.25	-3.61	2729
Bedford B.	2014-04-09	44.690	-63.640	60	31.07	1.60	272.41	0.160	7.58	1.24	11.38	3.07	0.32	-9.36	8173
Bedford B.	2014-04-16	44.690	-63.640	1	28.69	4.05	341.86	12.470	1.31	0.44	9.39	0.50	0.20	-2.83	26787
Bedford B.	2014-04-16	44.690	-63.640	5	28.74	3.98	334.18	16.710	1.39	0.45	9.41	0.46	0.21	-2.91	38057
Bedford B.	2014-04-16	44.690	-63.640	10	29.54	3.10	332.33	18.940	3.56	0.64	8.38	0.65	0.22	-3.78	6276
Bedford B.	2014-04-16	44.690	-63.640	60	31.03	1.59	273.66	0.140	7.55	1.33	12.07	3.41	0.31	-10.83	18288
Bedford B.	2014-04-23	44.690	-63.640	1	27.65	6.80	403.01	16.710	0.00	0.18	9.29	0.48	0.07	0.02	24706
Bedford B.	2014-04-23	44.690	-63.640	5	29.56	3.97	361.48	22.840	0.00	0.35	8.90	0.49	0.08	-2.70	11162
Bedford B.	2014-04-23	44.690	-63.640	10	30.34	1.94	318.20	8.140	0.00	0.35	6.10	0.68	0.08	-2.70	67838
Bedford B.	2014-04-23	44.690	-63.640	60	31.01	1.59	255.35	0.270	7.85	1.35	12.42	3.57	0.36	-10.85	6713
Bedford B.	2014-04-30	44.690	-63.640	1	27.71	4.10	340.27	1.230	1.19	0.32	7.66	2.75	0.10	-1.03	6894
Bedford B.	2014-04-30	44.690	-63.640	5	29.57	3.46	335.78	1.870	0.45	0.34	5.56	1.17	0.10	-2.09	30056
Bedford B.	2014-04-30	44.690	-63.640	10	30.31	2.25	327.59	1.610	1.81	0.52	4.02	1.63	0.16	-3.61	11333
Bedford B.	2014-04-30	44.690	-63.640	60	31.00	1.56	259.92	0.250	7.74	1.48	13.57	4.06	0.34	-13.04	113027
Bedford B.	2014-05-06	44.690	-63.640	1	28.67	5.39	322.83	2.080	1.14	0.33	5.55	2.39	0.11	-1.24	26214
Bedford B.	2014-05-06	44.690	-63.640	5	29.49	4.18	329.80	2.670	0.97	0.41	4.92	2.08	0.11	-2.69	4057
Bedford B.	2014-05-06	44.690	-63.640	10	30.26	2.55	319.63	0.810	1.78	0.53	3.86	2.64	0.14	-3.80	5992
Bedford B.	2014-05-06	44.690	-63.640	60	30.97	1.56	243.13	0.200	8.82	1.49	13.48	4.17	0.25	-12.12	54360
Bedford B.	2014-05-14	44.690	-63.640	1	28.63	7.91	347.61	4.710	0.00	0.19	1.01	0.45	0.07	-0.14	14850
Bedford B.	2014-05-14	44.690	-63.640	5	29.76	4.97	333.08	4.750	0.00	0.25	1.43	1.01	0.00	-1.10	24293
Bedford B.	2014-05-14	44.690	-63.640	10	30.16	3.43	321.25	3.950	0.76	0.49	2.70	1.54	0.10	-4.18	11363
Bedford B.	2014-05-14	44.690	-63.640	60	30.96	1.55	235.80	0.110	8.89	1.51	14.07	4.41	0.19	-12.37	22052
Bedford B.	2014-05-21	44.690	-63.640	1	29.07	9.77	326.05	2.120	0.00	0.17	0.00	0.81	0.04	0.18	7783
Bedford B.	2014-05-21	44.690	-63.640	5	29.70	7.57	346.00	1.950	0.00	0.19	0.21	0.64	0.00	-0.14	29014
Bedford B.	2014-05-21	44.690	-63.640	10	30.25	4.06	338.78	4.670	0.22	0.34	1.17	0.88	0.08	-2.32	6417
Bedford B.	2014-05-21	44.690	-63.640	60	30.99	1.57	223.78	0.070	9.20	1.67	15.73	5.92	0.19	-14.62	5053
Bedford B.	2014-05-28	44.690	-63.640	1	29.49	7.42	344.58	2.760	0.00	0.22	0.21	1.08	0.00	-0.62	994
Bedford B.	2014-05-28	44.690	-63.640	5	29.71	6.90	329.66	5.600	0.00	0.29	0.29	1.01	0.00	-1.74	645
Bedford B.	2014-05-28	44.690	-63.640	10	30.21	4.95	324.87	6.200	0.05	0.37	1.07	0.78	0.00	-2.97	657
Bedford B.	2014-05-28	44.690	-63.640	60	30.97	1.57	210.51	0.080	9.28	1.69	16.57	6.72	0.16	-14.86	484
Bedford B.	2014-06-04	44.690	-63.640	1	29.59	11.30	352.02	4.050	0.00	0.19	0.00	0.49	0.00	-0.14	795
Bedford B.	2014-06-04	44.690	-63.640	5	29.77	9.95	358.49	4.310	0.00	0.25	0.00	0.58	0.00	-1.10	653
Bedford B.	2014-06-04	44.690	-63.640	10	30.12	6.22	367.27	0.100	0.00	0.30	0.07	0.61	0.00	-1.90	725
Bedford B.	2014-06-04	44.690	-63.640	60	30.96	1.57	206.22	0.050	9.03	1.70	17.02	7.31	0.14	-15.27	897
Bedford B.	2014-06-11	44.690	-63.640	1	29.24	13.09	360.08	3.130	0.19	0.19	1.24	1.31	0.08	0.05	475
Bedford B.	2014-06-11	44.690	-63.640	5	29.89	9.69	393.44	4.640	0.00	0.24	0.51	0.51	0.06	-0.94	1662

Bedford B.	2014-06-11	44.690	-63.640	10	30.12	7.19	325.15	4.350	0.00	0.28	0.58	0.81	0.00	-1.58	1045
Bedford B.	2014-06-11	44.690	-63.640	60	30.94	1.57	199.10	0.140	9.11	1.78	17.52	7.68	0.12	-16.47	538
Bedford B.	2014-06-18	44.690	-63.640	5	29.67	11.13	323.67	2.360	0.04	0.23	1.30	0.95	0.06	-0.74	4645
Bedford B.	2014-06-18	44.690	-63.640	10	30.06	8.40	311.96	3.870	0.00	0.29	1.44	1.17	0.08	-1.74	682
Bedford B.	2014-06-18	44.690	-63.640	60	30.93	1.58	187.97	0.790	8.84	1.85	18.25	7.91	0.13	-17.86	1004
Bedford B.	2014-06-25	44.690	-63.640	1	28.94	14.09	295.05	3.790	0.33	0.17	2.42	0.84	0.11	0.51	594
Bedford B.	2014-06-25	44.690	-63.640	5	29.63	11.48	312.85	3.500	0.00	0.17	1.79	0.61	0.08	0.18	1488
Bedford B.	2014-06-25	44.690	-63.640	10	30.00	8.92	315.09	2.060	0.00	0.24	1.78	1.25	0.08	-0.94	1025
Bedford B.	2014-07-02	44.690	-63.640	1	27.72	17.98	320.74	2.540	0.00	0.15	0.00	1.02	0.14	0.50	1345
Bedford B.	2014-07-02	44.690	-63.640	5	29.74	10.25	361.96	4.790	0.00	0.16	0.00	0.61	0.07	0.34	1687
Bedford B.	2014-07-02	44.690	-63.640	10	30.03	8.80	298.39	0.130	0.00	0.25	0.14	0.97	0.10	-1.10	2919
Bedford B.	2014-07-02	44.690	-63.640	60	30.92	1.61	152.32	1.150	8.78	2.07	20.78	10.32	0.24	-21.44	2770
Bedford B.	2014-07-08	44.690	-63.640	1	28.88	14.02	292.67	3.090	0.00	0.15	0.00	0.52	0.05	0.50	476
Bedford B.	2014-07-08	44.690	-63.640	5	29.62	11.26	295.41	4.020	0.00	0.19	0.46	0.37	0.04	-0.14	174
Bedford B.	2014-07-08	44.690	-63.640	10	30.22	8.43	281.89	0.120	0.00	0.45	1.83	0.37	0.06	-4.30	924
Bedford B.	2014-07-08	44.690	-63.640	60	30.92	1.62	138.71	1.110	9.37	1.97	21.57	12.84	0.21	-19.25	1291
Bedford B.	2014-07-15	44.690	-63.640	1	29.11	16.52	286.67	12.630	0.55	0.23	0.10	1.10	0.12	-0.23	2572
Bedford B.	2014-07-15	44.690	-63.640	5	30.43	7.53	304.94	4.570	0.00	0.23	1.25	0.22	0.05	-0.78	883
Bedford B.	2014-07-15	44.690	-63.640	60	31.24	3.85	276.82	0.550	2.95	0.89	7.03	3.96	0.24	-8.39	2208
Bedford B.	2014-07-23	44.690	-63.640	1	29.23	15.80	286.95	2.800	0.00	0.09	0.59	0.53	0.07	1.46	454
Bedford B.	2014-07-23	44.690	-63.640	5	30.34	8.35	310.56	2.060	0.00	0.17	1.29	0.51	0.06	0.18	312
Bedford B.	2014-07-23	44.690	-63.640	10	30.71	5.33	281.57	3.650	0.00	0.52	4.76	0.59	0.10	-5.42	1180
Bedford B.	2014-07-23	44.690	-63.640	60	31.21	3.86	272.73	0.190	3.52	1.16	9.38	6.41	0.33	-12.14	646
Bedford B.	2014-07-30	44.690	-63.640	1	28.48	17.88	280.15	17.060	0.00	0.20	0.00	0.66	0.11	-0.30	731
Bedford B.	2014-07-30	44.690	-63.640	5	30.44	10.32	311.68	4.490	0.00	0.23	1.45	0.78	0.06	-0.78	484
Bedford B.	2014-07-30	44.690	-63.640	10	30.85	5.27	279.26	5.560	0.54	0.50	4.12	0.65	0.13	-4.56	185
Bedford B.	2014-07-30	44.690	-63.640	60	31.23	3.79	229.22	0.170	3.73	1.17	9.65	6.68	0.35	-12.09	79
Bedford B.	2014-08-06	44.690	-63.640	1	29.74	17.16	281.12	4.350	0.00	0.15	0.00	0.39	0.07	0.50	11428
Bedford B.	2014-08-06	44.690	-63.640	5	30.39	12.73	297.54	4.830	0.00	0.21	0.59	0.57	0.06	-0.46	235
Bedford B.	2014-08-06	44.690	-63.640	60	31.26	3.99	239.56	0.140	4.04	1.44	12.46	8.99	0.45	-16.10	133
Bedford B.	2014-08-13	44.690	-63.640	1	29.63	18.63	332.66	3.540	0.00	0.16	0.48	0.29	0.05	0.34	70
Bedford B.	2014-08-13	44.690	-63.640	5	30.49	13.52	317.01	4.310	0.00	0.23	1.07	0.53	0.04	-0.78	1564
Bedford B.	2014-08-13	44.690	-63.640	10	30.72	10.34	281.02	8.870	0.57	0.52	3.15	1.11	0.14	-4.85	714
Bedford B.	2014-08-13	44.690	-63.640	60	31.25	3.86	205.32	0.100	3.95	1.49	12.97	12.29	0.32	-16.99	2286
Bedford B.	2014-08-20	44.690	-63.640	1	30.42	15.19	311.51	2.650	0.00	0.22	1.16	0.74	0.03	-0.62	1417
Bedford B.	2014-08-20	44.690	-63.640	5	30.58	13.34	306.53	4.750	0.00	0.22	0.86	0.76	0.03	-0.62	1095
Bedford B.	2014-08-20	44.690	-63.640	10	30.88	9.41	273.84	5.340	0.00	0.40	1.16	1.19	0.10	-3.50	681
Bedford B.	2014-08-20	44.690	-63.640	60	31.24	3.85	187.51	0.100	4.74	1.69	15.71	13.28	0.16	-19.40	753
Bedford B.	2014-08-27	44.690	-63.640	1	29.89	17.69	277.77	2.910	0.04	0.15	1.40	0.66	0.06	0.54	2785
Bedford B.	2014-08-27	44.690	-63.640	5	30.04	17.80	271.78	3.390	0.00	0.18	1.52	0.59	0.08	0.02	1257

Bedford B.	2014-08-27	44.690	-63.640	10	30.44	15.90	246.45	7.850	0.05	0.39	2.12	0.56	0.09	-3.29	2134
Bedford B.	2014-08-27	44.690	-63.640	60	31.24	3.88	191.90	0.070	5.31	1.74	17.18	13.44	0.14	-19.63	1564
Bedford B.	2014-09-03	44.690	-63.640	1	30.01	19.20	280.04	2.540	0.00	0.19	1.64	0.48	0.07	-0.14	2513
Bedford B.	2014-09-03	44.690	-63.640	5	30.13	18.19	286.24	4.310	0.00	0.25	1.85	0.45	0.06	-1.10	4573
Bedford B.	2014-09-03	44.690	-63.640	10	30.39	15.86	214.43	2.060	1.55	0.72	4.30	4.19	0.29	-7.07	3608
Bedford B.	2014-09-03	44.690	-63.640	60	31.25	3.86	167.54	0.050	6.11	1.79	17.48	13.22	0.16	-19.63	1992
Bedford B.	2014-09-10	44.690	-63.640	1	30.13	18.19	285.73	4.490	0.21	0.40	2.36	2.03	0.07	-3.29	2126
Bedford B.	2014-09-10	44.690	-63.640	5	30.59	13.34	232.24	3.280	0.68	0.45	3.26	1.97	0.11	-3.62	27701
Bedford B.	2014-09-10	44.690	-63.640	10	30.87	10.34	244.62	1.660	2.23	0.77	5.39	4.51	0.25	-7.19	2396
Bedford B.	2014-09-10	44.690	-63.640	60	31.23	3.87	162.72	0.070	7.11	1.70	17.66	12.22	0.16	-17.19	2569
Bedford B.	2014-09-17	44.690	-63.640	1	30.37	14.73	270.70	8.530	0.00	0.23	0.41	0.68	0.06	-0.78	60390
Bedford B.	2014-09-17	44.690	-63.640	5	30.86	11.35	257.50	11.940	0.00	0.27	0.23	0.56	0.06	-1.42	30922
Bedford B.	2014-09-17	44.690	-63.640	10	31.06	9.39	247.26	13.820	0.08	0.39	0.68	0.76	0.06	-3.26	18819
Bedford B.	2014-09-17	44.690	-63.640	60	31.25	3.89	152.61	0.090	8.55	1.94	20.51	13.03	0.19	-19.59	2436
Bedford B.	2014-09-24	44.690	-63.640	1	30.68	15.02	324.05	6.700	0.85	0.40	1.41	2.39	0.09	-2.65	23223
Bedford B.	2014-09-24	44.690	-63.640	5	30.69	14.18	304.31	3.090	0.63	0.40	1.31	3.17	0.08	-2.87	9706
Bedford B.	2014-10-01	44.690	-63.640	1	29.03	15.49	275.09	4.900	1.44	0.43	2.04	3.88	0.12	-2.54	21462
Bedford B.	2014-10-01	44.690	-63.640	5	29.61	15.53	261.09	1.150	2.55	0.75	4.35	5.44	0.27	-6.55	35458
Bedford B.	2014-09-24	44.690	-63.640	10	30.71	13.60	280.10	8.020	0.83	0.36	1.31	2.02	0.10	-2.03	6873
Bedford B.	2014-10-01	44.690	-63.640	60	31.24	3.92	130.20	0.090	12.16	2.00	20.72	6.94	0.21	-16.94	8744
Bedford B.	2014-10-08	44.690	-63.640	1	28.88	15.36	248.56	4.460	0.00	0.18	0.00	0.73	0.07	0.02	1810
Bedford B.	2014-10-08	44.690	-63.640	5	30.58	11.52	247.73	16.380	0.39	0.31	0.06	1.18	0.08	-1.67	798
Bedford B.	2014-10-08	44.690	-63.640	10	31.07	9.25	241.97	13.310	1.47	0.47	0.52	2.99	0.15	-3.15	1610
Bedford B.	2014-10-08	44.690	-63.640	60	31.24	3.94	122.79	0.240	13.88	2.24	23.60	6.47	0.15	-19.06	777
Bedford B.	2014-10-15	44.690	-63.640	1	29.73	14.16	318.19	2.360	0.44	0.12	0.19	1.23	0.09	1.42	1551
Bedford B.	2014-10-15	44.690	-63.640	5	30.15	13.88	273.18	3.680	0.56	0.17	0.29	1.15	0.10	0.74	786
Bedford B.	2014-10-15	44.690	-63.640	10	30.62	12.36	251.41	2.650	0.92	0.16	0.23	1.98	0.12	1.26	5285
Bedford B.	2014-10-15	44.690	-63.640	60	31.23	3.97	104.11	0.290	16.04	2.20	23.81	4.54	0.13	-16.26	3236
Bedford B.	2014-10-22	44.690	-63.640	1	30.11	13.37	293.69	5.230	1.29	0.31	1.28	2.86	0.16	-0.77	450
Bedford B.	2014-10-22	44.690	-63.640	5	30.15	13.23	286.79	5.820	1.30	0.32	1.23	3.01	0.15	-0.92	64395
Bedford B.	2014-10-22	44.690	-63.640	10	30.37	12.86	255.71	2.210	1.93	0.42	1.48	4.12	0.19	-1.89	415
Bedford B.	2014-10-22	44.690	-63.640	60	31.22	4.01	98.17	0.150	16.86	2.60	26.01	4.10	0.11	-21.84	188
Bedford B.	2014-10-29	44.690	-63.640	1	29.81	12.91	272.99	2.580	3.35	0.55	5.44	5.96	0.25	-2.55	466
Bedford B.	2014-10-29	44.690	-63.640	5	29.84	12.92	270.21	3.610	2.98	0.56	4.10	5.70	0.24	-3.08	150
Bedford B.	2014-10-29	44.690	-63.640	10	30.35	12.83	230.98	1.020	3.01	0.70	3.73	6.17	0.31	-5.29	70701
Bedford B.	2014-10-29	44.690	-63.640	60	31.23	4.07	88.43	0.100	16.66	3.12	32.13	4.62	0.14	-30.36	365
Bedford B.	2014-11-05	44.690	-63.640	1	28.68	11.65	238.16	14.670	3.80	0.55	2.11	3.64	0.31	-2.10	691
Bedford B.	2014-11-05	44.690	-63.640	5	29.44	12.06	245.36	12.970	3.45	0.69	3.03	4.78	0.32	-4.69	652
Bedford B.	2014-11-05	44.690	-63.640	10	30.38	12.28	221.58	4.570	3.43	0.88	3.86	8.79	0.35	-7.75	536
Bedford B.	2014-11-05	44.690	-63.640	60	31.23	4.14	77.89	0.180	18.51	2.33	23.05	1.44	0.16	-15.87	151

Bedford B.	2014-11-12	44.690	-63.640	1	28.52	10.44	263.43	10.410	4.20	0.28	1.13	2.23	0.42	2.62	1113
Bedford B.	2014-11-12	44.690	-63.640	5	29.35	11.76	243.07	5.530	3.98	0.46	0.88	3.16	0.36	-0.48	39310
Bedford B.	2014-11-12	44.690	-63.640	10	30.22	11.99	218.22	2.060	4.38	0.83	4.63	5.10	0.46	-6.00	516
Bedford B.	2014-11-12	44.690	-63.640	60	31.23	4.20	74.74	0.260	20.10	2.58	30.92	0.66	0.28	-18.28	598
Bedford B.	2014-11-19	44.690	-63.640	1	27.75	10.40	283.78	1.230	6.55	0.81	6.76	6.51	0.43	-3.51	6095
Bedford B.	2014-11-19	44.690	-63.640	5	29.73	11.69	226.28	1.470	6.19	0.80	6.07	5.96	0.41	-3.71	472
Bedford B.	2014-11-19	44.690	-63.640	10	30.58	11.46	207.11	1.430	5.94	0.81	4.88	4.78	0.42	-4.12	670
Bedford B.	2014-11-19	44.690	-63.640	60	31.23	4.23	51.10	0.100	20.09	2.53	28.04	0.57	0.08	-17.49	750
Bedford B.	2014-11-26	44.690	-63.640	1	28.52	8.59	268.87	1.690	7.01	0.83	10.17	6.21	0.45	-3.37	2392
Bedford B.	2014-11-26	44.690	-63.640	5	29.11	8.97	262.49	1.230	7.13	0.85	7.13	4.50	0.44	-3.57	1547
Bedford B.	2014-11-26	44.690	-63.640	10	30.39	10.13	212.85	0.800	8.00	0.91	7.34	2.84	0.44	-3.66	388
Bedford B.	2014-11-26	44.690	-63.640	60	31.24	4.38	55.73	0.070	19.33	2.98	32.47	0.51	0.24	-25.45	148
Bedford B.	2014-12-03	44.690	-63.640	1	28.49	9.13	233.74	2.390	9.10	0.96	10.13	3.31	0.48	-3.36	471
Bedford B.	2014-12-03	44.690	-63.640	5	30.07	9.29	238.18	2.070	8.90	0.97	10.30	3.43	0.47	-3.72	552
Bedford B.	2014-12-03	44.690	-63.640	10	30.70	8.99	231.07	1.140	9.21	1.02	9.28	2.22	0.47	-4.21	17287
Bedford B.	2014-12-03	44.690	-63.640	60	31.24	4.49	51.57	0.240	19.10	2.66	30.63	0.43	0.08	-20.56	702
Bedford B.	2014-17-10	44.690	-63.640	60	31.26	4.67	48.69	NA	18.72	2.59	31.32	0.45	0.09	-19.82	1194
Bedford B.	2014-17-10	44.690	-63.640	60	31.27	4.90	57.04	NA	18.42	2.58	32.44	0.36	0.05	-19.96	660
GEOVIDE	1	40.333	-10.036	6	35.11	16.71	250.80	0.164	0.03		0.90		0.00		250081
GEOVIDE	1	40.333	-10.036	11	35.14	16.61	253.10	0.185	0.02		0.87		0.00		41026
GEOVIDE	1	40.333	-10.036	25	35.27	16.00	260.70	0.326							20419
GEOVIDE	1	40.333	-10.036	34	35.31	15.63	260.80	0.351	0.86		0.55		0.00		5655
GEOVIDE	1	40.333	-10.036	51	35.68	13.77	255.10	0.461	4.09		1.09		0.00		6124
GEOVIDE	1	40.333	-10.036	65	35.68	13.81	248.70	0.432							4194
GEOVIDE	1	40.333	-10.036	80	35.71	13.45	252.40	0.175	5.42		1.88		0.00		3695
GEOVIDE	1	40.333	-10.036	170	35.71	12.65	236.50		6.11		2.17		0.00		5583
GEOVIDE	1	40.333	-10.036	200	35.70	12.58	232.80	0.009	7.06		2.72		0.00		4497
GEOVIDE	2	40.333	-9.459	11	34.96	16.88	252.20	0.208	0.03		0.93		0.02		5008
GEOVIDE	2	40.333	-9.459	35	35.58	14.52	257.70	0.981	0.95		0.37		0.12		8991
GEOVIDE	2	40.333	-9.459	100	35.79	13.33	238.70	0.036	5.29		1.49		0.03		9157
GEOVIDE	2	40.333	-9.459	141	35.76	12.99	221.70	0.024	7.45		3.73		0.06		5752
GEOVIDE	4	40.333	-9.767	8	35.15	16.31	253.10	0.280	0.02		0.85		0.01		5754
GEOVIDE	4	40.333	-9.767	35	35.66	14.11	264.10	0.777	0.93		0.31		0.26		6663
GEOVIDE	4	40.333	-9.767	100	35.74	13.05	240.00	0.043	5.87		1.74		0.03		6449
GEOVIDE	4	40.333	-9.767	199	35.70	12.50	230.20	0.005	8.20		2.74		0.02		7340
GEOVIDE	11	40.333	-12.219	6	35.83	15.90	252.30	0.253	0.01		0.10		0.00		2801
GEOVIDE	11	40.333	-12.219	25	35.82	16.01	253.70	0.334							2854
GEOVIDE	11	40.333	-12.219	54	35.81	13.70	257.10	0.586	5.05		1.42		0.04		13521
GEOVIDE	11	40.333	-12.219	200	35.74	12.82	243.50	0.004	6.82		2.07		0.01		5534
GEOVIDE	13	41.384	-13.888	5	35.84	15.48	256.70	0.537	0.04		0.53		0.01		3222
GEOVIDE	13	41.384	-13.888	25	35.85	15.43	257.30		0.62		0.67		0.06		14754
GEOVIDE	13	41.384	-13.888	35	35.77	13.77	263.40	0.675							7833

GEOVIDE	13	41.384	-13.888	76	35.76	13.16	250.10	0.169				8215
GEOVIDE	13	41.384	-13.888	850	35.64	10.06	181.00		16.54	8.69	0.00	6087
GEOVIDE	15	42.000	-15.461	5	35.69	14.83	259.90	0.470	0.76	0.38	0.08	5595
GEOVIDE	15	42.000	-15.461	20	35.72	14.91	261.70	0.542				13260
GEOVIDE	15	42.000	-15.461	50	35.79	14.82	259.80	0.489	4.86	1.80	0.62	9013
GEOVIDE	15	42.000	-15.461	200	35.66	12.15	248.40	0.002	8.08	2.71	0.00	10496
GEOVIDE	17	43.781	-17.323	1	35.68	14.62	260.20	0.335	1.50	0.87	0.09	9081
GEOVIDE	17	43.781	-17.323	15	35.69	14.44	260.80	0.482	1.69	0.86	0.16	9779
GEOVIDE	17	43.781	-17.323	60	35.67	12.56	253.20	0.232				6692
GEOVIDE	17	43.781	-17.323	187	35.64	11.98	247.00	0.006	8.44	2.82	0.01	11294
GEOVIDE	19	45.550	-18.500	5	35.68	15.04	264.60	0.304	0.97	1.00	0.10	1538
GEOVIDE	19	45.550	-18.500	25	35.68	13.93	265.10	0.287				12798
GEOVIDE	19	45.550	-18.500	50	35.66	12.82	260.00	0.355	1.42	1.26	0.15	13354
GEOVIDE	19	45.550	-18.500	197	35.65	12.07	252.60	0.007	8.56	3.19	0.01	17188
GEOVIDE	21	46.543	-19.672	3	35.69	14.48	277.80	0.779	0.77	0.40	0.09	3067
GEOVIDE	21	46.543	-19.672	15	35.67	14.09	277.10	1.208				608
GEOVIDE	21	46.543	-19.672	30	35.66	13.52	277.50	0.734	3.93	1.62	0.15	691
GEOVIDE	21	46.543	-19.672	70	35.63	12.62	262.60	0.059				10085 5
GEOVIDE	21	46.543	-19.672	397	35.55	11.38	246.10		11.13	4.54	0.01	2141
GEOVIDE	21	46.543	-19.672	800	35.30	8.82	188.30		17.54	10.05	0.00	43
GEOVIDE	23	48.039	-20.847	6	35.68	13.84	272.40	1.040	1.49	0.19	0.15	15472
GEOVIDE	23	48.039	-20.847	9	35.69	13.86	270.40	1.052				912
GEOVIDE	23	48.039	-20.847	20	35.68	13.84	269.50	0.990	1.53	0.43	0.16	942
GEOVIDE	23	48.039	-20.847	51	35.71	13.24	253.80	0.274				6235
GEOVIDE	23	48.039	-20.847	200	35.64	12.10	248.40	0.007	8.61	3.34	0.02	12827
GEOVIDE	25	49.529	-22.172	4	35.58	12.93	274.60	0.736				98441
GEOVIDE	25	49.529	-22.172	9	35.58	12.93	274.60	0.737	3.46	1.17	0.20	3044
GEOVIDE	25	49.529	-22.172	56	35.57	11.98	263.50	0.176				48858
GEOVIDE	25	49.529	-22.172	200	35.52	11.11	252.40	0.007	9.71	3.88	0.03	7155
GEOVIDE	26	50.278	-22.602	7	35.37	11.92	279.40	0.690	5.42	1.13	0.53	7946
GEOVIDE	26	50.278	-22.602	25	35.37	11.92	280.30	0.707				1932
GEOVIDE	26	50.278	-22.602	35	35.36	11.57	278.90	0.697	5.68	1.19	0.55	4434
GEOVIDE	26	50.278	-22.602	70	35.30	10.39	273.00	0.430				1359
GEOVIDE	26	50.278	-22.602	500	34.98	6.56	210.20		17.79	11.26	0.01	2804
GEOVIDE	29	53.020	-24.752	5	34.99	10.16	296.70	1.143	6.25	1.63	0.16	11063
GEOVIDE	29	53.020	-24.752	20	35.02	10.17	295.40	1.217	7.25	2.18	0.17	1903
GEOVIDE	29	53.020	-24.752	54	35.04	8.49	293.70	0.608				4425
GEOVIDE	29	53.020	-24.752	201	34.95	7.09	262.30	0.007	14.28	7.88	0.13	1467
GEOVIDE	30	54.000	-25.533	400	34.84	5.23	251.30		15.64	9.26	0.01	1136
GEOVIDE	32	55.500	-26.711	5	35.12	10.39	289.80	1.041				1230
GEOVIDE	32	55.500	-26.711	20	35.13	10.48	290.60	1.099	6.53	1.07	0.15	1546
GEOVIDE	32	55.500	-26.711	25	35.12	10.33	289.20	1.051				337
GEOVIDE	32	55.500	-26.711	70	35.12	8.70	289.90	0.327				783
GEOVIDE	32	55.500	-26.711	450	35.06	6.46	218.40		17.24	10.65	0.01	6123

GEOVIDE	34	57.004	-28.879	4	35.17	10.51	288.90	0.548	6.26	0.80	0.11	664
GEOVIDE	34	57.004	-28.879	10	35.18	10.47	289.50	0.581				300
GEOVIDE	34	57.004	-28.879	20	35.17	10.15	290.30	0.664	6.58	0.93	0.17	531
GEOVIDE	34	57.004	-28.879	50	35.16	8.76	288.50	0.362				1012
GEOVIDE	34	57.004	-28.879	60	35.15	8.61	284.50	0.226				980
GEOVIDE	34	57.004	-28.879	193	35.10	7.68	264.50	0.035	13.94	6.90	0.04	534
GEOVIDE	36	58.207	-29.725	3	35.05	9.53	291.90	0.432	6.70	0.41	0.14	723
GEOVIDE	36	58.207	-29.725	10	35.05	9.53	292.00	0.422				423
GEOVIDE	36	58.207	-29.725	20	35.05	9.48	294.90	0.476	7.10	0.70	0.16	502
GEOVIDE	36	58.207	-29.725	30	35.04	8.41	300.70	0.612				348
GEOVIDE	36	58.207	-29.725	40	35.04	7.98	300.30	0.652				151
GEOVIDE	36	58.207	-29.725	50	35.04	7.88	295.20	0.617				436
GEOVIDE	36	58.207	-29.725	200	35.04	6.74	278.50	0.029	13.52	6.86	0.02	162
GEOVIDE	38	58.843	-31.267	5	35.06	9.38	295.50	0.805	6.38	0.51	0.14	501
GEOVIDE	38	58.843	-31.267	10	35.06	9.37	295.60	0.744	6.36	0.61	0.14	391
GEOVIDE	38	58.843	-31.267	15	35.06	9.37	295.60	0.719	6.39	0.52	0.14	748
GEOVIDE	38	58.843	-31.267	29	35.06	9.33	295.50	0.707	6.72	0.64	0.15	65
GEOVIDE	38	58.843	-31.267	50	35.07	8.23	295.00	0.468	7.93	1.21	0.40	97
GEOVIDE	38	58.843	-31.267	60	35.10	7.95	291.20	0.249	8.57	1.75	0.68	17
GEOVIDE	38	58.843	-31.267	200	35.15	7.56	271.80	0.055	13.13	6.44	0.03	25
GEOVIDE	38	58.843	-31.267	650	35.08	6.03	234.80		16.39	10.42	0.01	152
GEOVIDE	40	59.102	-33.829	7	34.98	8.43	307.10	0.968	6.41	1.71	0.17	175
GEOVIDE	40	59.102	-33.829	10	34.98	8.41	307.00	1.014				46
GEOVIDE	40	59.102	-33.829	20	34.98	8.38	307.50	1.136				256
GEOVIDE	40	59.102	-33.829	30	35.00	7.71	310.30	2.077				223
GEOVIDE	40	59.102	-33.829	42	34.97	7.12	308.30	4.317	8.87	5.88	0.26	140
GEOVIDE	40	59.102	-33.829	50	34.99	7.02	300.00	1.519				1682
GEOVIDE	40	59.102	-33.829	61	35.01	6.95	290.10					74
GEOVIDE	40	59.102	-33.829	70	35.01	6.80	287.50	0.345				976
GEOVIDE	40	59.102	-33.829	200	35.01	6.16	281.30	0.060	14.00	7.14	0.01	416
GEOVIDE	42	59.363	-36.396	5	34.96	7.83	309.30	1.141	7.29	2.72	0.19	175
GEOVIDE	42	59.363	-36.396	10	34.96	7.83	309.30	1.210				388
GEOVIDE	42	59.363	-36.396	22	34.96	7.83	309.00	0.989				120
GEOVIDE	42	59.363	-36.396	30	34.92	6.82	307.90	1.805	7.65	3.68	0.19	142
GEOVIDE	42	59.363	-36.396	40	34.93	6.54	305.30	1.008				413
GEOVIDE	42	59.363	-36.396	50	34.94	6.23	300.50	0.471				303
GEOVIDE	42	59.363	-36.396	60	34.94	5.93	296.50					50
GEOVIDE	42	59.363	-36.396	70	34.95	5.76	293.30	0.191				158
GEOVIDE	42	59.363	-36.396	200	34.92	4.77	286.10	0.025	15.05	8.32	0.02	137
GEOVIDE	44	59.623	-38.954	5	34.85	6.84	318.40	1.920				437
GEOVIDE	44	59.623	-38.954	10	34.85	6.88	316.50	1.487	9.09	7.75	0.11	307
GEOVIDE	44	59.623	-38.954	20	34.85	6.79	317.70	1.880	14.84	8.10	0.05	248
GEOVIDE	44	59.623	-38.954	70	34.90	4.44	296.70	0.025				303
GEOVIDE	49	59.773	-49.297	5	34.88	6.65	313.50	1.807	8.46	5.91	0.12	877

GEOVIDE	49	59.773	-49.297	10	34.88	6.63	313.60	1.952				403
GEOVIDE	49	59.773	-49.297	20	34.88	6.60	312.70	1.894				1053
GEOVIDE	49	59.773	-49.297	30	34.89	6.55	310.70	1.818				447
GEOVIDE	49	59.773	-49.297	40	34.91	6.55	311.80	0.889	8.69	5.41	0.14	494
GEOVIDE	49	59.773	-49.297	50	34.98	6.17	304.80	0.751				464
GEOVIDE	49	59.773	-49.297	60	34.98	6.16	302.50					1255
GEOVIDE	49	59.773	-49.297	80	34.99	5.85	298.60		12.35			166
GEOVIDE	49	59.773	-49.297	200	34.97	5.18	288.40	0.095	14.08	7.62	0.09	248
GEOVIDE	53	59.902	-43.015	5	31.89	-0.71	423.50	3.677	0.00	2.00	0.00	270
GEOVIDE	53	59.902	-43.015	10	31.66	-1.21	431.70	3.533				228
GEOVIDE	53	59.902	-43.015	25	32.08	-1.05	423.60	4.291	0.00	2.42	0.00	725
GEOVIDE	53	59.902	-43.015	30	32.34	-0.86	417.30	4.928				190
GEOVIDE	53	59.902	-43.015	75	32.84	-1.41	366.30	2.292				33
GEOVIDE	53	59.902	-43.015	160	33.37	-0.63	350.90	0.227	6.43	5.13	0.10	129
GEOVIDE	56	59.823	-42.399	5	34.12	2.34	327.00	0.330				20
GEOVIDE	56	59.823	-42.399	10	34.24	2.81	326.00	0.316	6.86	4.64	0.28	193
GEOVIDE	56	59.823	-42.399	20	34.39	3.44	323.60	0.305	7.64	4.39	0.28	132
GEOVIDE	56	59.823	-42.399	30	34.52	4.06	321.10	0.289				345
GEOVIDE	56	59.823	-42.399	40	34.62	4.69	319.00	0.658	7.96	4.08	0.32	741
GEOVIDE	56	59.823	-42.399	60	34.73	5.33	316.40	1.057	8.21	4.94	0.33	229
GEOVIDE	56	59.823	-42.399	70	34.70	5.08	316.60	0.324				62
GEOVIDE	60	59.799	-42.013	5	34.90	6.90	325.20	2.484	7.42	4.29	0.11	819
GEOVIDE	60	59.799	-42.013	20	34.90	6.69	318.50	2.993				494
GEOVIDE	60	59.799	-42.013	30	34.88	6.68	312.70	2.606				225
GEOVIDE	60	59.799	-42.013	70	35.02	6.22	297.10	0.270				675
GEOVIDE	61	59.753	-45.112	5	32.45	-0.07	410.50	4.424	0.01	2.86	0.00	1004
GEOVIDE	61	59.753	-45.112	10	32.40	0.04	412.70	5.822				870
GEOVIDE	61	59.753	-45.112	20	32.57	-0.23	407.50	6.650				905
GEOVIDE	61	59.753	-45.112	30	32.58	-0.23	404.70	6.212				935
GEOVIDE	61	59.753	-45.112	40	32.73	-0.48	394.60	5.645				755
GEOVIDE	61	59.753	-45.112	50	32.81	-0.63	383.40	3.892	2.05	3.79	0.04	615
GEOVIDE	61	59.753	-45.112	65	32.85	-0.59	372.60	3.316				1183
GEOVIDE	61	59.753	-45.112	80	32.96	-1.07	365.10	1.272				1473
GEOVIDE	61	59.753	-45.112	100	33.11	-1.29	356.90	0.368	6.69	5.29	0.08	548
GEOVIDE	61	59.753	-45.112	160	33.56	-0.04	344.60		7.91	5.56	0.10	447
GEOVIDE	63	59.434	-45.689	5	34.63	5.26	314.00	0.398	9.34	4.23	0.15	426
GEOVIDE	63	59.434	-45.689	10	34.44	3.91	315.70	0.392				450
GEOVIDE	63	59.434	-45.689	20	34.71	5.53	313.10	0.396				320
GEOVIDE	63	59.434	-45.689	30	34.75	5.65	313.90	0.387				279
GEOVIDE	63	59.434	-45.689	40	34.69	4.87	313.80	0.283				402
GEOVIDE	63	59.434	-45.689	50	34.66	4.61	313.40	0.227	10.62	5.25	0.18	234
GEOVIDE	63	59.434	-45.689	60	34.66	4.52	312.80					416
GEOVIDE	63	59.434	-45.689	70	34.69	4.43	311.30	0.128				1033
GEOVIDE	63	59.434	-45.689	80	34.74	4.63	309.50					584

GEOVIDE	63	59.434	-45.689	100	34.79	4.72	303.80	0.103	12.30	6.17	0.20	494
GEOVIDE	63	59.434	-45.689	150	34.85	4.86	299.00	0.071	13.21	6.54	0.21	953
GEOVIDE	63	59.434	-45.689	200	34.88	4.89	297.00	0.048	14.00	6.92	0.20	400
GEOVIDE	64	59.068	-46.083	15	34.70	6.68	327.60	0.401				1652
GEOVIDE	64	59.068	-46.083	30	34.74	6.14	331.90	0.445				2015
GEOVIDE	64	59.068	-46.083	70	34.88	5.12	303.80	0.111				611
GEOVIDE	68	56.913	-47.419	3	34.52	6.25	327.00	0.475	0.02	2.24	0.00	640
GEOVIDE	68	56.913	-47.419	10	34.53	6.15	328.70	0.848				669
GEOVIDE	68	56.913	-47.419	20	34.51	5.41	336.30	1.108				518
GEOVIDE	68	56.913	-47.419	30	34.54	4.80	338.30	1.087				1046
GEOVIDE	68	56.913	-47.419	40	34.60	4.36	328.00	1.096				548
GEOVIDE	68	56.913	-47.419	50	34.69	3.66	308.80	0.229	8.79	6.65	0.21	587
GEOVIDE	68	56.913	-47.419	60	34.71	3.63	303.30	0.093				610
GEOVIDE	68	56.913	-47.419	70	34.72	3.54	298.90					602
GEOVIDE	68	56.913	-47.419	80	34.76	3.67	295.70	0.044				593
GEOVIDE	68	56.913	-47.419	100	34.82	3.88	294.70	0.033	14.05	7.16	0.40	831
GEOVIDE	68	56.913	-47.419	150	34.84	3.87	297.20	0.021	14.32	7.27	0.09	465
GEOVIDE	68	56.913	-47.419	200	34.85	3.86	296.90		14.53	7.46	0.05	469
GEOVIDE	69	55.841	-48.094	5	34.60	6.27	326.10	0.372	0.07	3.56	0.00	8685
GEOVIDE	69	55.841	-48.094	20	34.61	6.15	329.40	1.003				5277
GEOVIDE	69	55.841	-48.094	25	34.66	4.41	322.50	0.458				27820
GEOVIDE	69	55.841	-48.094	70	34.82	3.92	293.30					3223
GEOVIDE	69	55.841	-48.094	1200	34.86	3.48	295.20		14.60	8.25	0.00	32806
GEOVIDE	71	53.692	-49.433	4	34.68	7.36	321.40	1.942	1.79	1.42	0.15	759
GEOVIDE	71	53.692	-49.433	10	34.68	7.36	321.60	2.212				268
GEOVIDE	71	53.692	-49.433	20	34.68	7.17	321.20	2.362				512
GEOVIDE	71	53.692	-49.433	30	34.68	6.88	318.60	2.158				33
GEOVIDE	71	53.692	-49.433	40	34.72	6.00	318.60	1.324				320
GEOVIDE	71	53.692	-49.433	50	34.72	5.78	310.90	1.222	8.23	6.33	0.00	100
GEOVIDE	71	53.692	-49.433	60	34.73	5.17	311.60					132
GEOVIDE	71	53.692	-49.433	70	34.73	4.46	304.10	0.075				94
GEOVIDE	71	53.692	-49.433	80	34.73	4.25	303.00					393
GEOVIDE	71	53.692	-49.433	100	34.75	4.06	300.90	0.019	14.57	7.73	0.15	71
GEOVIDE	71	53.692	-49.433	150	34.74	3.65	302.70	0.011	13.56	7.71	0.00	123
GEOVIDE	71	53.692	-49.433	200	34.78	3.72	296.20	0.007				9
GEOVIDE	77	52.995	-51.096	3	34.47	7.39	325.10	2.069	0.33	1.09	0.04	119
GEOVIDE	77	52.995	-51.096	17	34.63	7.28	317.70	2.475	2.89	3.59	0.13	399
GEOVIDE	77	52.995	-51.096	30	34.65	4.83	332.80	1.940				341
GEOVIDE	77	52.995	-51.096	70	34.76	3.46	307.30					266
GEOVIDE	78	51.989	-53.820	5	31.81	5.37	338.40	0.183	0.05	0.20	0.00	32
GEOVIDE	78	51.989	-53.820	20	32.80	-0.14	404.10	0.108				193
GEOVIDE	78	51.989	-53.820	30	32.92	-0.87	426.90	9.568				1406
GEOVIDE	78	51.989	-53.820	35	32.96	-1.21	407.90	4.905				110
GEOVIDE	78	51.989	-53.820	45	33.04	-1.37	346.20	0.559	7.30	8.80	0.10	118

GEOVIDE	78	51.989	-53.820	60	33.11	-1.47	337.20	0.192				342
GEOVIDE	78	51.989	-53.820	100	33.36	-1.20	329.30		9.05	8.50	0.07	496
GEOVIDE	78	51.989	-53.820	150	33.65	-0.64	324.20	0.091	10.24	8.40	0.06	0
GEOVIDE	78	51.989	-53.820	200	34.06	0.67	316.70	0.093	11.64	8.60	0.13	50

Supplemental Table 11: Custom designed *nifH* primers with Illumina adaptors.

Name	Fusion sequence (5'-3')
nifHF1-S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S513	AATGATACGGCGACCACCGAGATCTACACTCGACTAGTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S515	AATGATACGGCGACCACCGAGATCTACACTTCTAGCTTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S516	AATGATACGGCGACCACCGAGATCTACACCCTAGAGTTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S517	AATGATACGGCGACCACCGAGATCTACACGCGTAAGATCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S518	AATGATACGGCGACCACCGAGATCTACACCTATTAAGTCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHF1-S522	AATGATACGGCGACCACCGAGATCTACACTTATGCGATCGTCGGC AGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA
nifHR2-N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC
nifHR2-N729	CAAGCAGAAGACGGCATACGAGATGACGTCGAGTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC

Supplemental Table 12: Read counts obtained from *nifH* high-throughput sequencing in the Atlantic Ocean.

	Max[1)]	Min[2)]	Mean	St. Dev.[3)]
AZMP HUD2014004/30	17076	4282	9222	2823
Bedford Basin	12877	7838	9650	1554
Discovery 361	50570	22	24601	18387
GEOTRACES Kn199/204	9676	1776	5660	1937
GEOVIDE	17504	0	4663	4620
Meteor 116	11466	35	3759	1483
Polarstern ANTXXVI-I	11311	3239	6542	1820

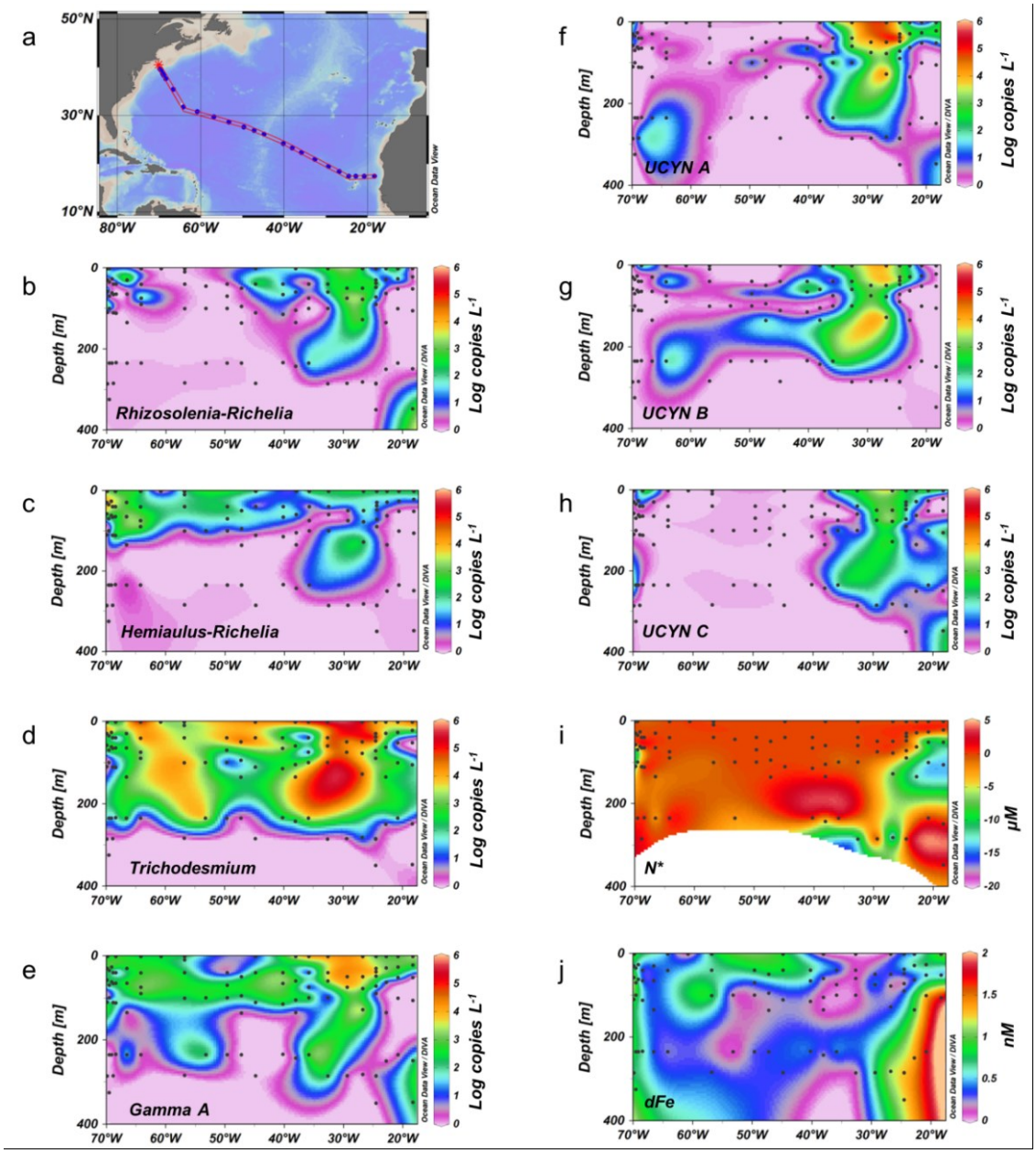
1) Maximum number of reads in a sample.
2) Maximum number of reads in a sample.
3) Standard Deviation of the mean number of reads.

Supplemental Table 13: Green gene reference number and genbank accession number of the 16S rRNA gene sequence of the most commonly[1)] detected OTUs in the Bedford Basin in 2014.

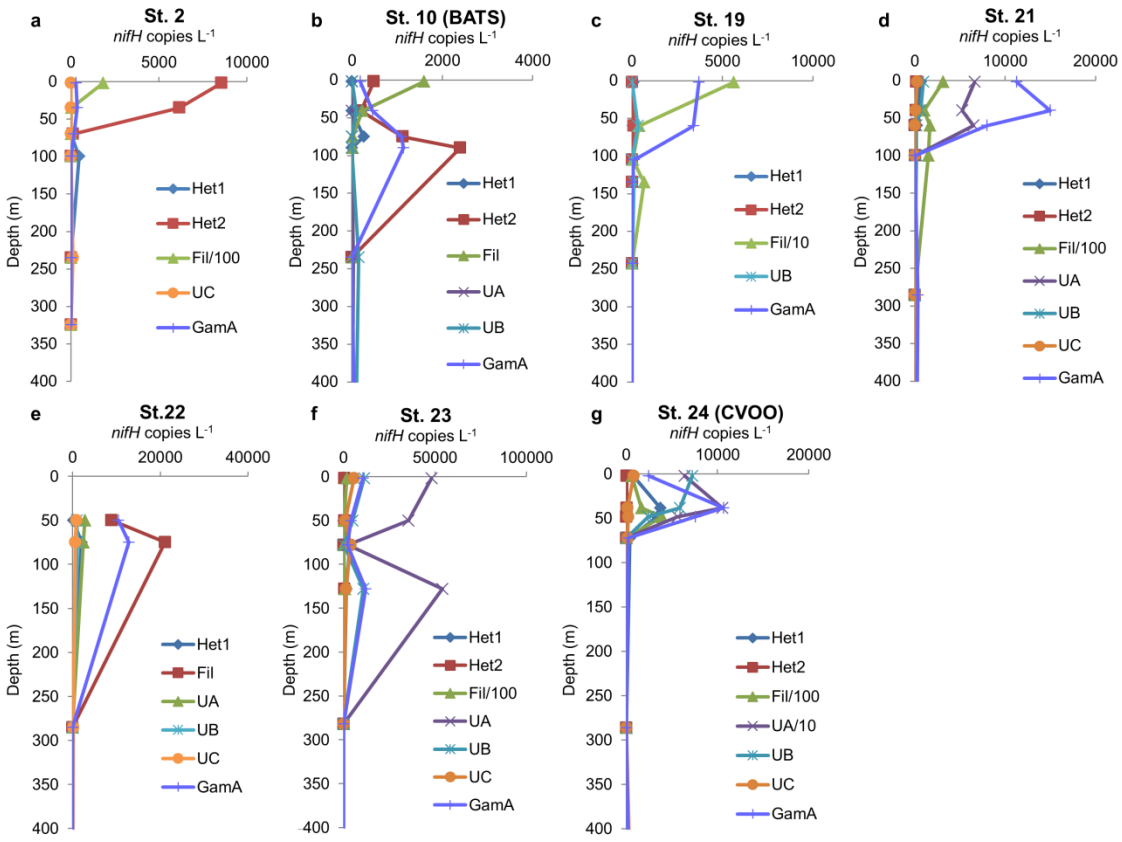
Name	Phylum	Class	Order	Family	Genus	gi number	prokMSA id
Colwellia	Proteobacteria	gamma-Proteobacteria	Alteromonadales	Colwelliaceae	*Colwellia*	NR_024805	644016
FlavobacteriaceaeA	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae		FR686298	795842
FlavobacteriaceaeB	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae		EU799118	311600
FlavobacteriaceaeC	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae		FR686275	659486
HTCC2207	Proteobacteria	gamma-Proteobacteria	Alteromonadales	Alteromonadaceae	*HTCC2207*	GQ348090	535135
Octadecabacter	Proteobacteria	alpha-Proteobacteria	Rhodobacterales	Rhodobacteraceae	*Octadecabacter*	FR683806	804483
PelagibacteraceaeA	Proteobacteria	alpha-Proteobacteria	Rickettsiales	Pelagibacteraceae		FR685476	637092
PelagibacteraceaeB	Proteobacteria	alpha-Proteobacteria	Rickettsiales	Pelagibacteraceae		EU800435	307744
Phaeobacter	Proteobacteria	alpha-Proteobacteria	Rhodobacterales	Rhodobacteraceae	*Phaeobacter*	FR683537	696544
PolaribacterA	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Polaribacter*	GQ349687	586650
PolaribacterB	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Polaribacter*	AF354621	28929
Pseudoalteromonas	Proteobacteria	gamma-Proteobacteria	Vibrionales	Pseudoalteromonadaceae	*Pseudoalteromonas*	HQ448943	827726
Psychrobacter	Proteobacteria	gamma-Proteobacteria	Pseudomonadales	Moraxellaceae	*Psychrobacter*	GQ443088	580411
RhodobacteraceaeA	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		FR685988	645011
RhodobacteraceaeB	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		FR685697	804449
RhodobacteraceaeC	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		FJ545507	534690
RhodobacteraceaeD	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		FJ826087	714708
RhodobacteraceaeE	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		FR686210	785501
RhodobacteraceaeF	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		EF659447	243160
RhodobacteraceaeG	Proteobacteria	gamma-Proteobacteria	Rhodobacterales	Rhodobacteraceae		EU544714	272142
SUP05	Proteobacteria	gamma-Proteobacteria	Oceanospirillales	SUP05		FJ628207	583880
UlvibacterA	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Ulvibacter*	FJ826059	774258
UlvibacterB	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Ulvibacter*	GQ452895	543487
ZA3409c	Actinobacteria	Acidimicrobiia	Acidimicrobiia	Acidimicrobiales	*ZA3409c*	GU474887	814290

1) OTUs that made up more than 1% of reads in each depth (1, 5, 10 and 60 m)

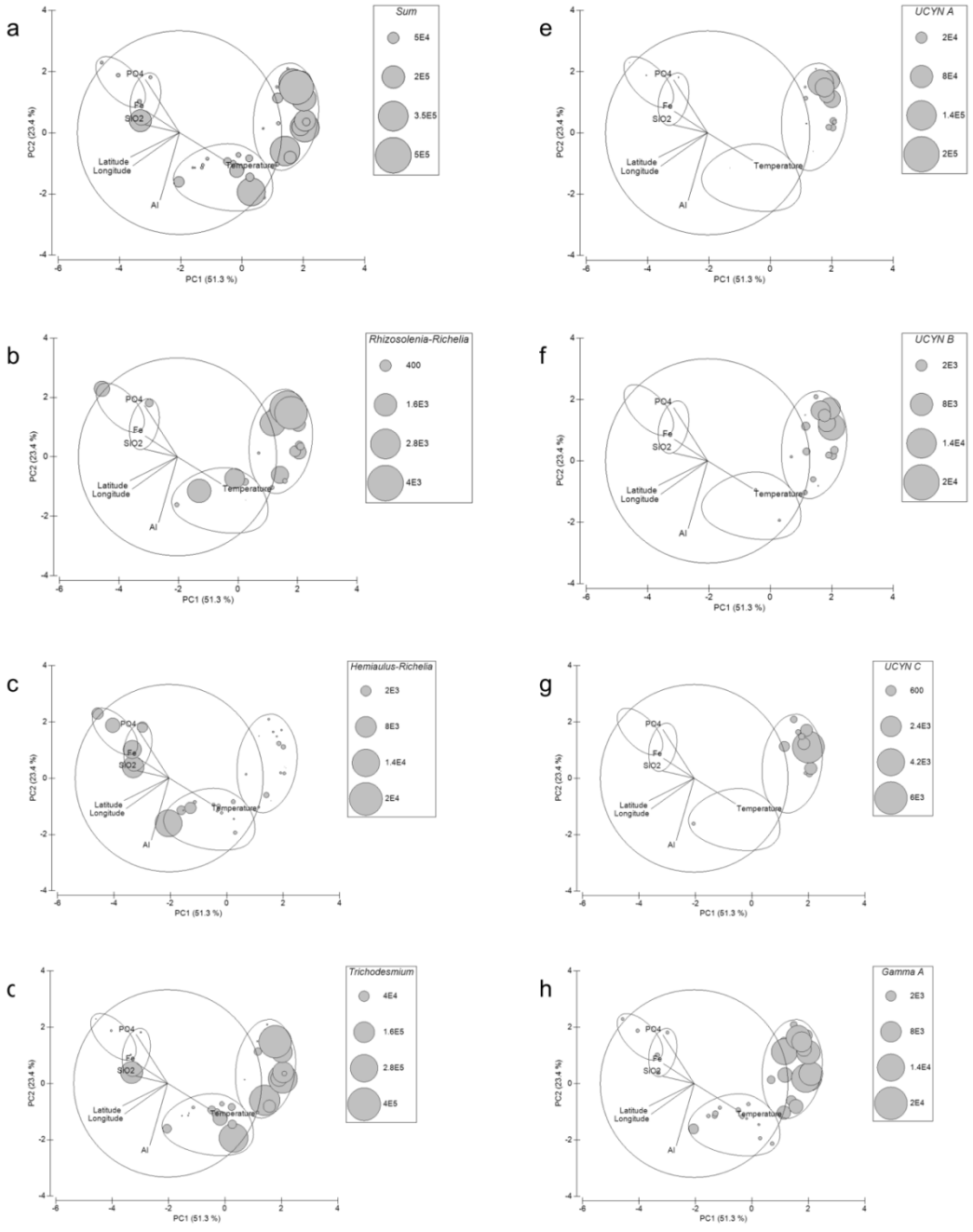
APPENDIX B: SUPPLEMENTAL FIGURES



Supplemental Figure 1: Section plots of stations across the west-east transect for *nifH* phylotypes, N* and dissolved Fe: a) Definition of section, b) *Rhizosolenia-Richelia* symbiont, c) *Hemiaulus-Richelia* symbiont, d) *Trichodesmium,* e) Gamma A, f) UCYN A, g) UCYN B, h) UCYN C, i) N* (N* = N – 16 P; Gruber and Sarmiento 1997), j) dissolved Fe (nM)

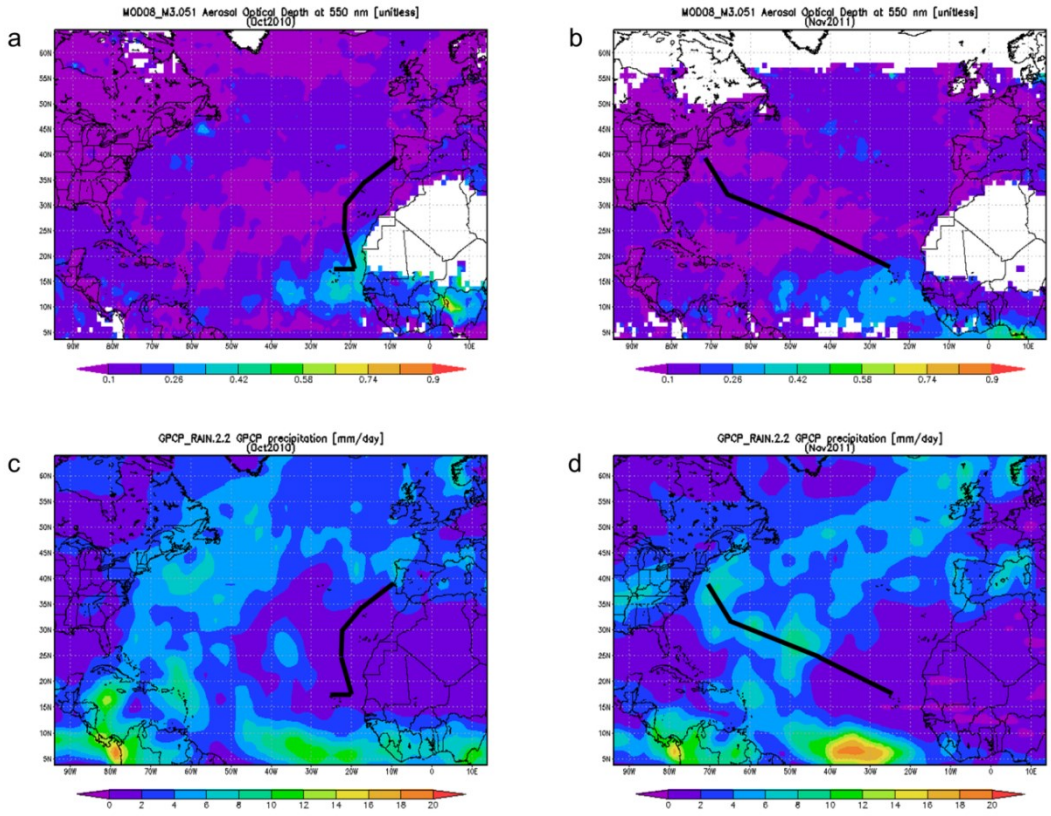


Supplemental Figure 2: Vertical profiles of *nifH* phylotypes L^{-1} at stations a) 2, b) 10, c) 19, d) 21, e) 22, f) 23, g) 24. If one phylotype was 10 or 100 fold higher that the others, this phylotype's abundance was divided by 10 or 100 as indicated in the ledgend.



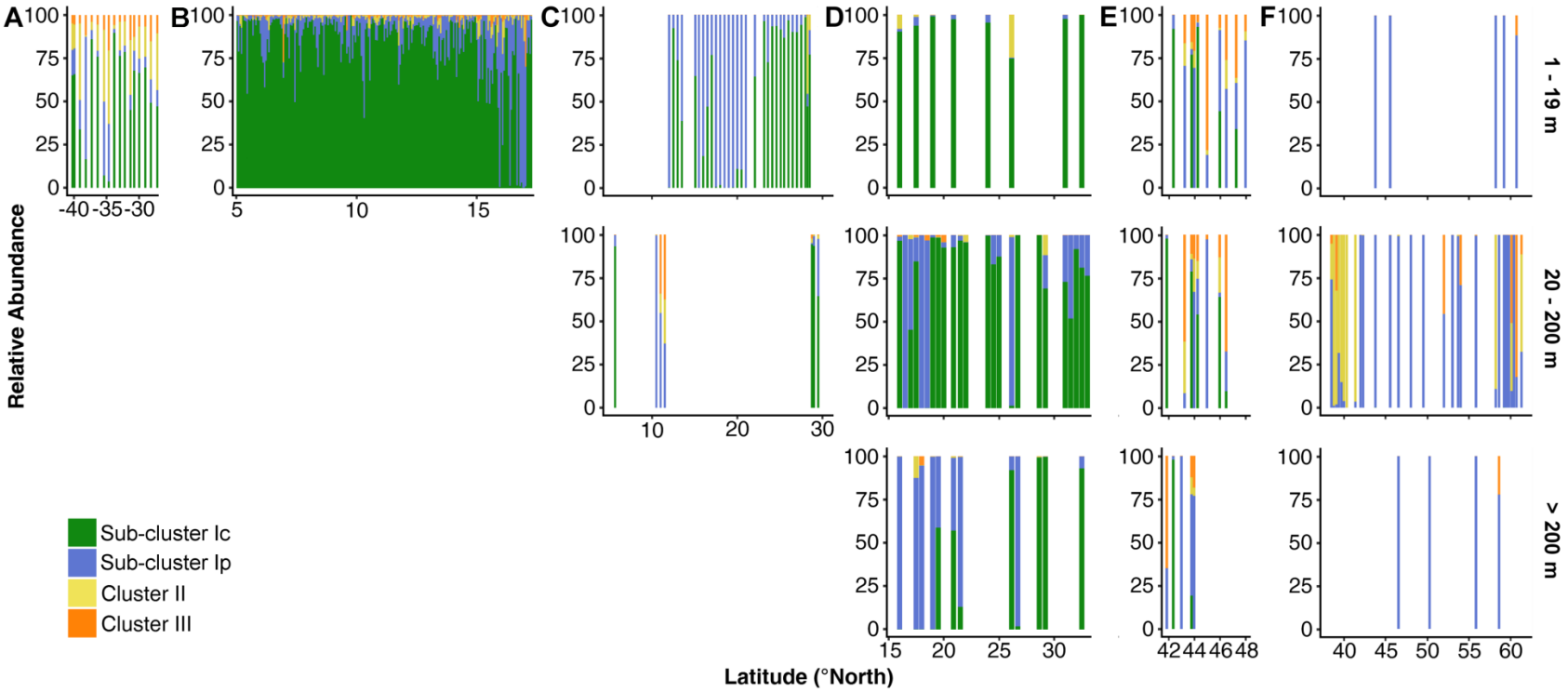
Supplemental Figure 3: Principal Components Analysis of SML samples from USGT10 and USGT11 showing variables that contributed to significant clustering of samples.

When two variables were highly correlated with each other only one was retained for the analysis. Significant clusters are traced with a line representing a Euclidean-distance of 3. The original PCA plot (Figure 2.5) was overlaid with abundances of a) *Rhizosolenia-Richelia* symbiont, b) *Hemiaulus-Richelia* symbiont, c) *Trichodesmium*, d) UCYN A, e) UCYN B, f) UCYN C and g) Gamma A.



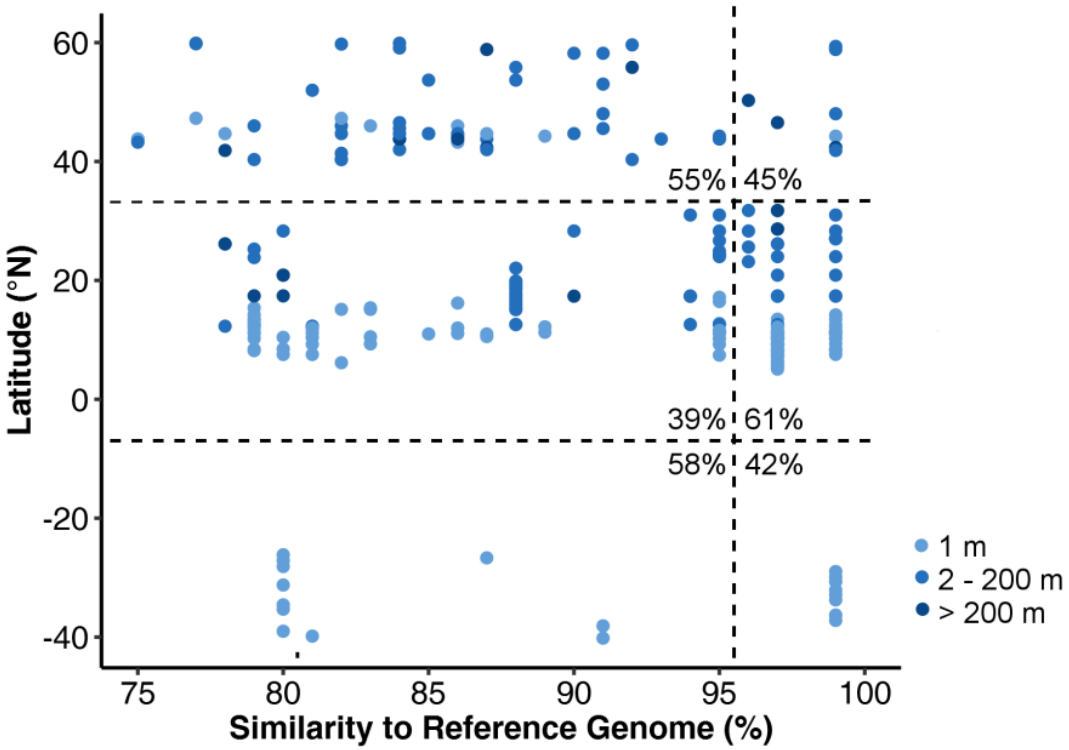
Supplemental Figure 4: Climatological conditions across the North Atlantic ocean during USGT10/11.

a) and b) are the aerosol dust deposition for Oct 2010 and Nov 2011, respectively. c) and d) are the average rain fall during Oct 2010 and Nov 2011, respectively.



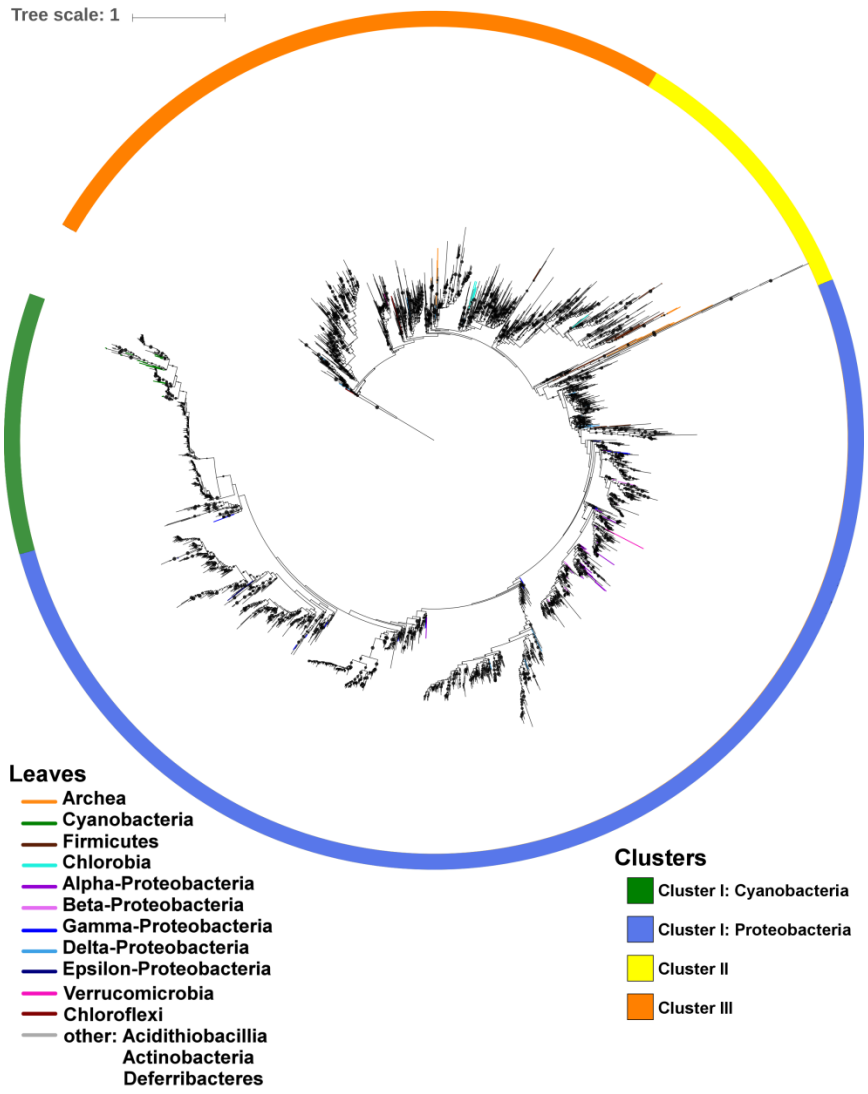
Supplemental Figure 5: Distribution of *nifH* clusters throughout the water column and along latitude.

Relative abundances of the major *nifH* clusters (green – cluster I: cyanobacteria, blue – cluster I: proteobacteria, yellow – cluster II, red – cluster III) is shown for every sample. Samples are listed as part of their cruise transect (A – Polarstern ANTXXVI-I, B - Meteor 116, C - Discovery D361, D - US-GEOTRACES Kn199 and Kn204, E - AZMP HUD2014004 and HUD2014030, F - GEOVIDE), according to their latitude and divided into 3 depths (1 m, 2 – 200 m and > 200 m).



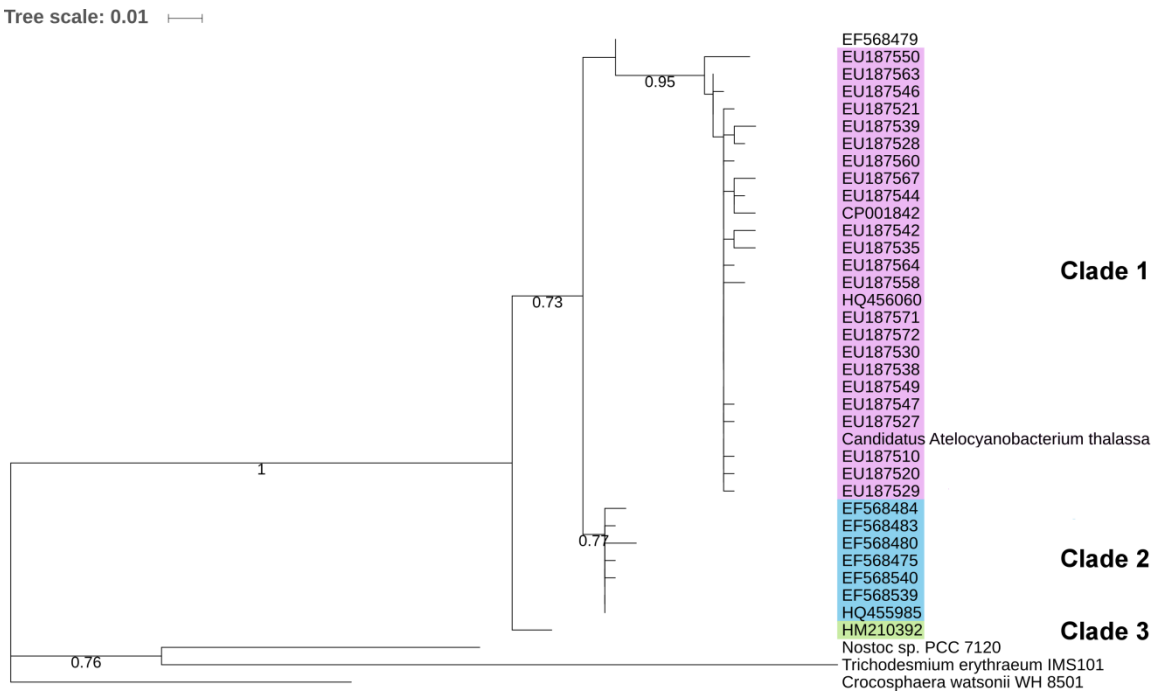
Supplemental Figure 6: Similarity of most commonly detected *nifH* OTU in each sample to the reference genome database.

The similarity of the most frequent OTU of every sample was recorded with respect to its most similar reverence genome sequence.



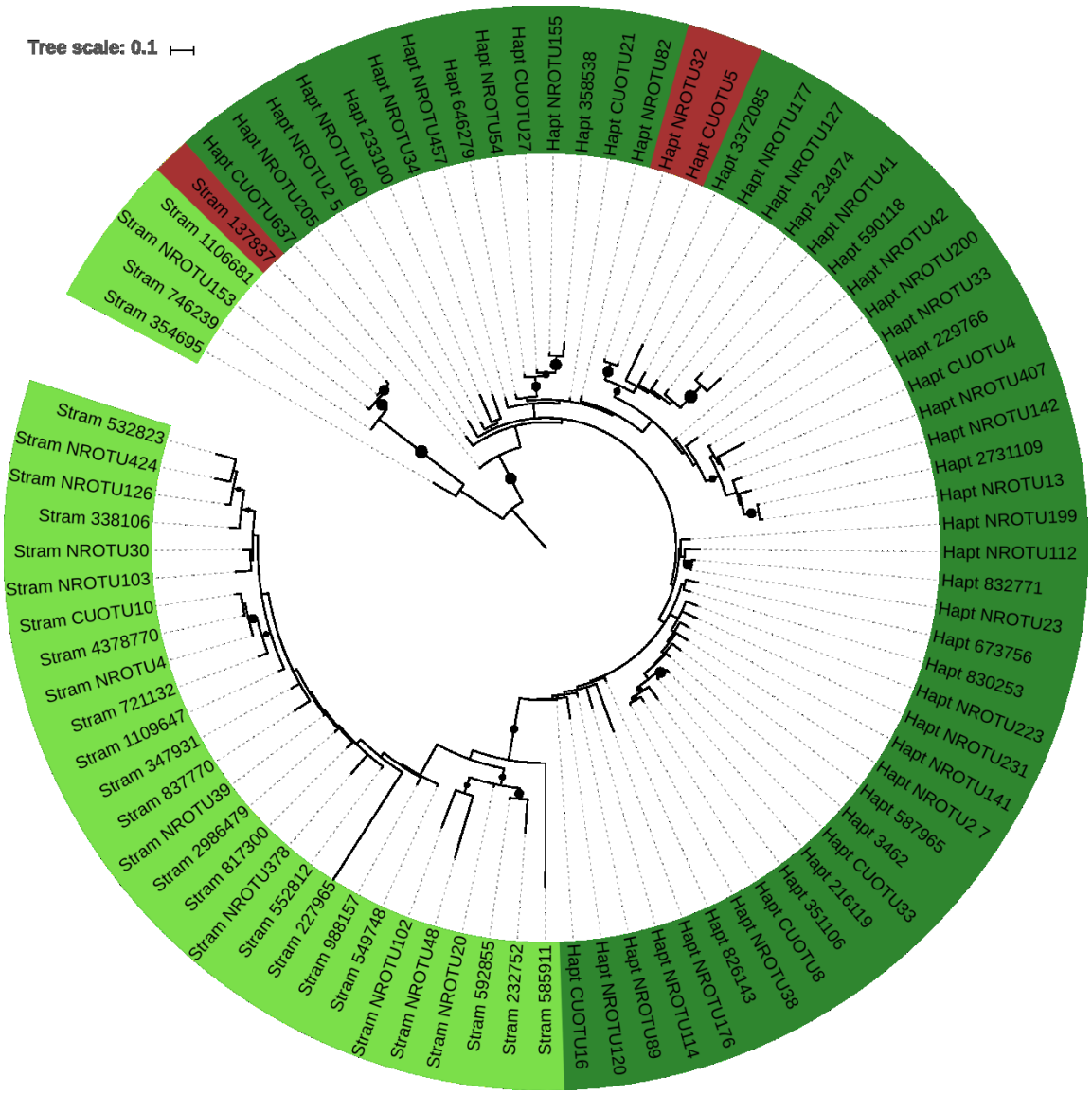
Supplemental Figure 7: Phylogenetic tree and abundance data for *nifH* sequences recovered throughout the Atlantic Ocean.

2655 OTUs were recovered through high-throughput sequencing of the *nifH* gene throughout the Atlantic Ocean. The phylogenetic affiliation of these OTUs and reference sequences was inferred by using the maximum likelihood method based on the GTR-GAMMA model (Stamakatis 2014). Bootstrap support from 100 replicates are indicated by circles when values were greater 50. The tree was displayed with branch lengths showing the number of substitutions per site. Leaves of sequences from reference genomes are coloured according to their taxonomy, while black leaves indicate environmental sequences. Clusters are assigned according to Zehr et al. (2003).



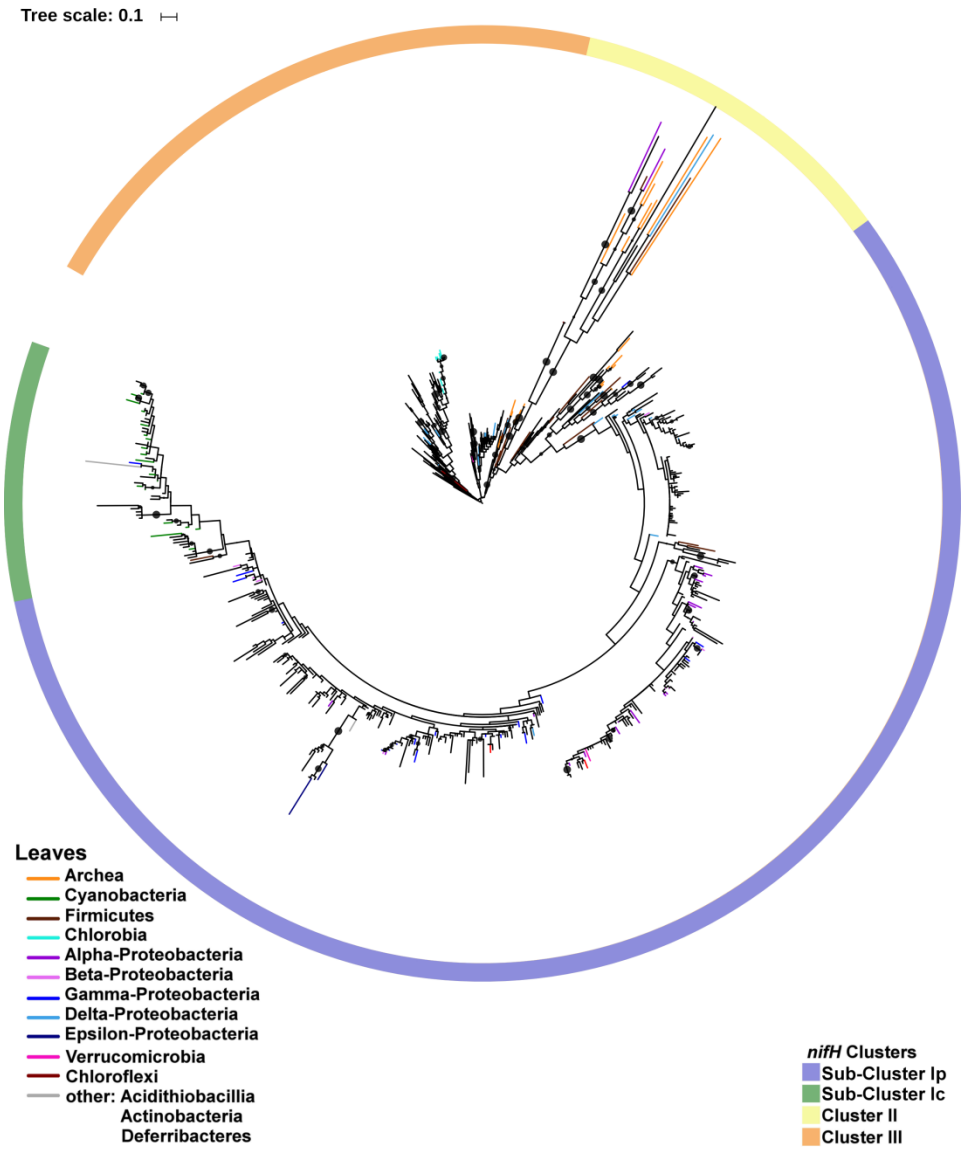
Supplemental Figure 8: Phylogenetic tree of *Candidatus* Atelocyanobacterium thalassa sequences.

At 96% sequence identity, 35 *Candidatus* A. thalassa OTUs were recovered through high-throughput sequencing of the *nifH* gene throughout the Atlantic Ocean. The evolutionary history of these OTUs was inferred using the maximum likelihood method based on the GTR-GAMMA model (Stamakatis 2014). The tree was displayed with branch lengths showing the number of substitutions per site. Bootstrapping values were calculated from 100 replicate trees. Values >70% are indicated on the according branches. Three clades of *Candidatus* A. thalassa (clade 1:purple, clade 2: blue, clade 3: green) were annotated according to Thompson et al. (2014) and *nifH* reference genome sequences are labelled.



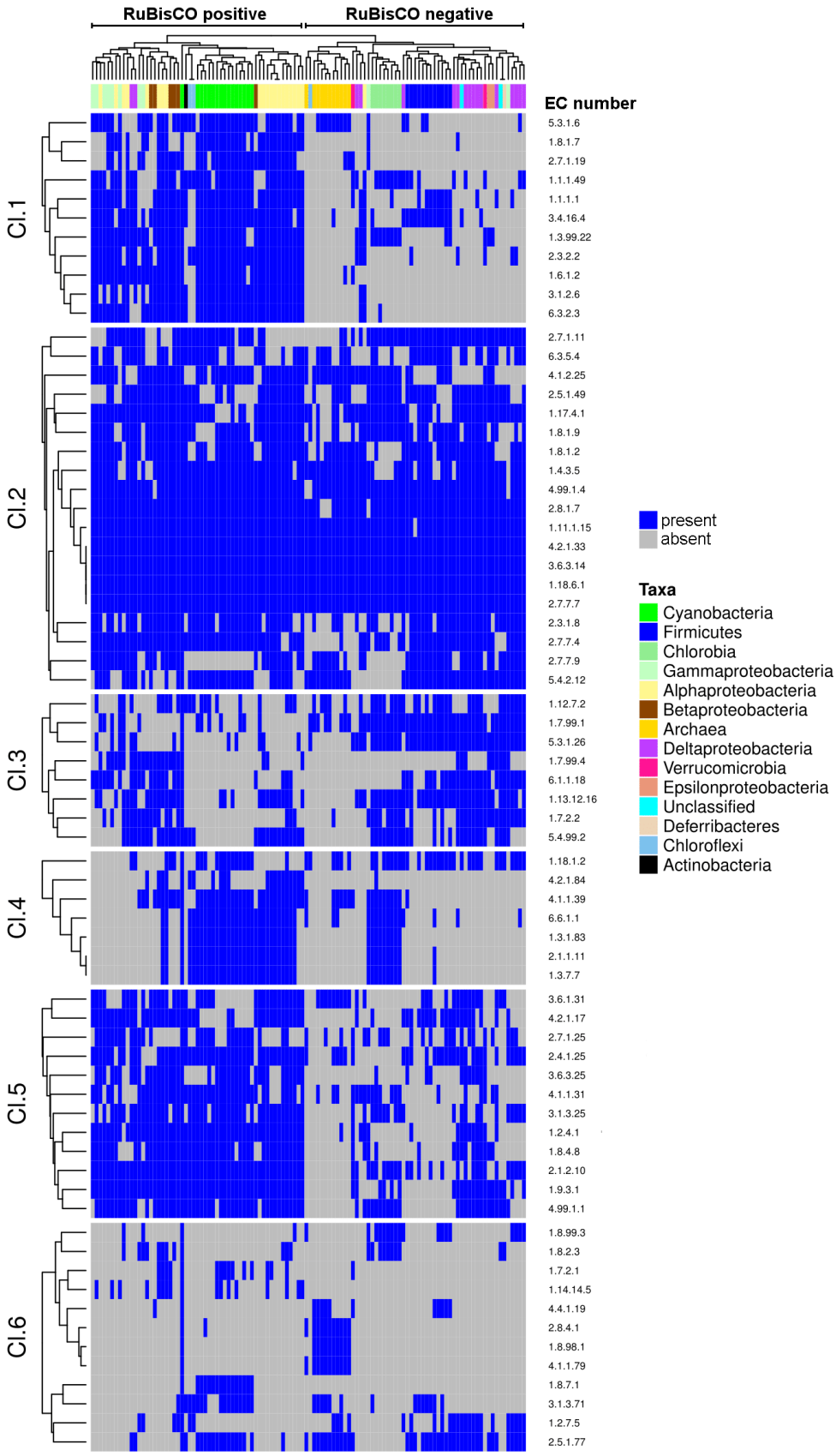
Supplemental Figure 9: Phylogenetic tree of chloroplasts 16S rDNA genes associated with *Candidatus* A. thalassa sequences.

After SparCC correlation, 83 chloroplast 16S rDNA genes were significantly associated with *Candidatus* A. thalassa sequences. Significant positive correlations are highlighted in red boxes. Correlations occurred between *Candidatus* A. thalassa and the taxa of Stramenopiles (light green) and Haptophyceae (dark green). The evolutionary history of these OTUs was inferred using the maximum likelihood method based on the GTR-GAMMA model (Stamakatis 2014). Bootstrapping values were calculated from 100 replicate trees. Values >50% are indicated as dots. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site and bootstrap values are indicated by black circles.



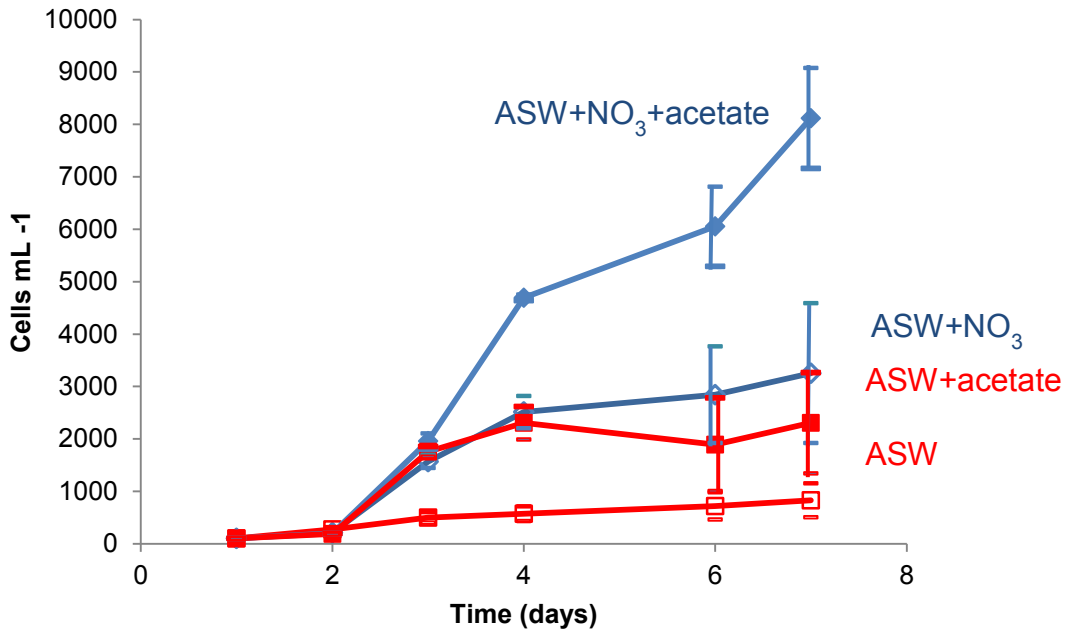
Supplemental Figure 10: Phylogenetic analysis based on protein sequences from extracted and clustered *nifH* sequences.

The *nifH* sequences from studies that identified non-cyanobacterial diazotrophs were extracted from NCBI along with their most closely related reference genomes. The sequences were clustered at 96% identity using CD-HIT (Li and Godzik 2006). The phylogenetic affiliation of their protein sequences was inferred by maximum likelihood tree building based on the WAG-GAMMA model of aligned sequences (MAFFT v. 7; Yamada et al. 2016; Stamakatis 2014; Katoh et al. 2002). Bootstrap values were calculated from 100 tree replicates and values >50% are shown as dots. The tree was displayed with branch lengths showing the number of substitutions per site Leaves of reference genomes are coloured according to their taxonomy. Clusters are assigned according to Zehr et al. (2003).


RuBisCO positive
RuBisCO negative
EC number
Cl.1
Cl.2
Cl.3
Cl.4
Cl.5
Cl.6
5.3.1.6
1.8.1.7
2.7.1.19
1.1.1.49
1.1.1.1
3.4.16.4
1.3.99.22
2.3.2.2
1.6.1.2
3.1.2.6
6.3.2.3
2.7.1.11
6.3.5.4
4.1.2.25
2.5.1.49
1.17.4.1
1.8.1.9
1.8.1.2
1.4.3.5
4.99.1.4
2.8.1.7
1.11.1.15
4.2.1.33
3.6.3.14
1.18.6.1
2.7.7.7
2.3.1.8
2.7.7.4
2.7.7.9
5.4.2.12
1.12.7.2
1.7.99.1
5.3.1.26
1.7.99.4
6.1.1.18
1.13.12.16
1.7.2.2
5.4.99.2
1.18.1.2
4.2.1.84
4.1.1.39
6.6.1.1
1.3.1.83
2.1.1.11
1.3.7.7
3.6.1.31
4.2.1.17
2.7.1.25
2.4.1.25
3.6.3.25
4.1.1.31
3.1.3.25
1.2.4.1
1.8.4.8
2.1.2.10
1.9.3.1
4.99.1.1
1.8.99.3
1.8.2.3
1.7.2.1
1.14.14.5
4.4.1.19
2.8.4.1
1.8.98.1
4.1.1.79
1.8.7.1
3.1.3.71
1.2.7.5
2.5.1.77
present
absent
Taxa
Cyanobacteria
Firmicutes
Chlorobia
Gammaproteobacteria
Alphaproteobacteria
Betaproteobacteria
Archaea
Deltaproteobacteria
Verrucomicrobia
Epsilonproteobacteria
Unclassified
Deferribacteres
Chloroflexi
Actinobacteria

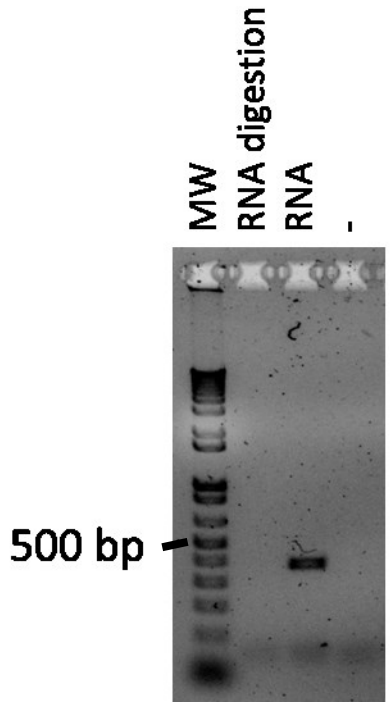
Supplemental Figure 11: Metabolic distinction between heterotrophic and autotrophic diazotrophs.

Enzymes distinguishing diazotrophic taxa were determined using PRIMER (Clarke and Gorley 2006), then, enzymes and diazotrophic reference genomes were clustered in *gplots* using R (R core team 2015; Warnes et al. 2013). EC numbers of enzymes are given along the vertical axis. Distinguishing features between RuBisCO positive and RuBisCO negative reference genomes are detected in Cl. I.



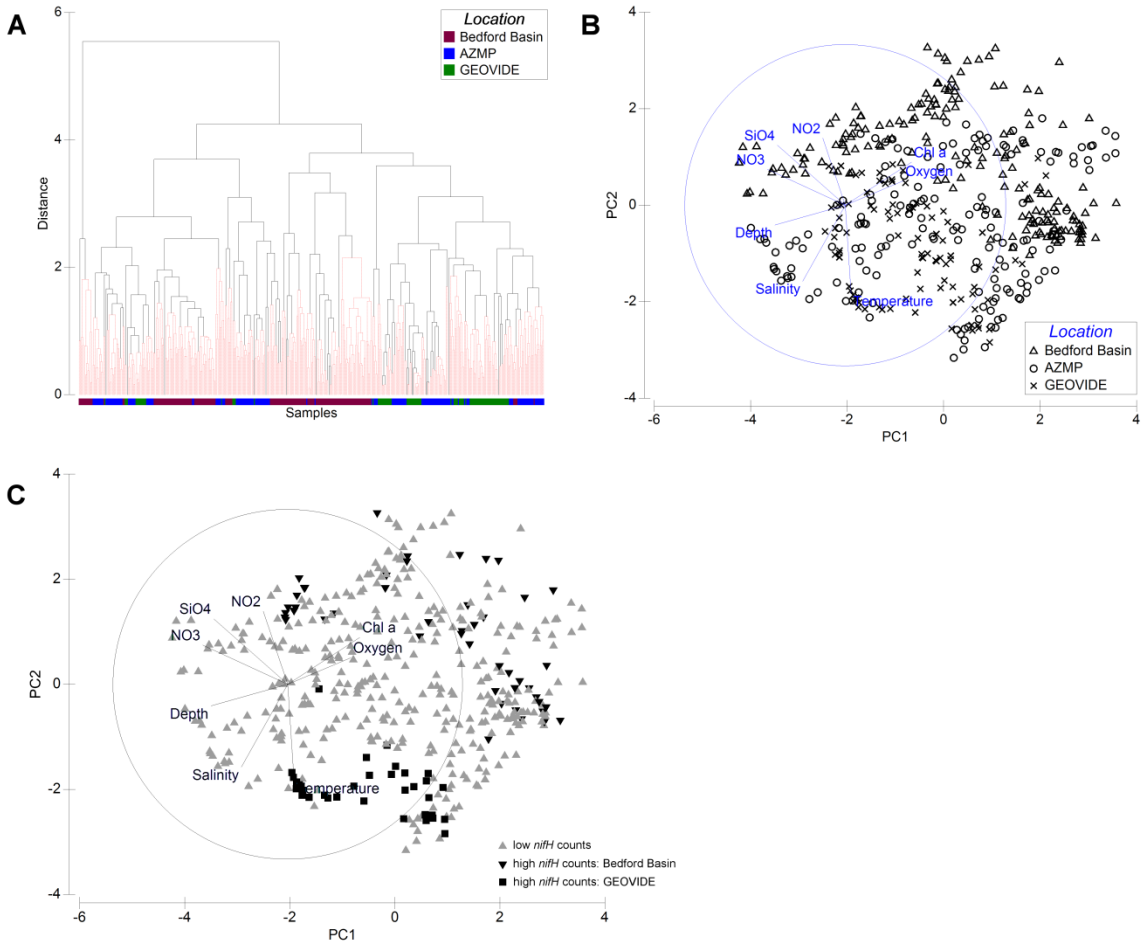
Supplemental Figure 12: Growth curves of isolated diazotroph.

The isolated diazotroph was grown in Artificial Sea water supplemented with acetate, nitrate or acetate + nitrate. Growth curves are shown for the priod of seven days post additionof nutrients.



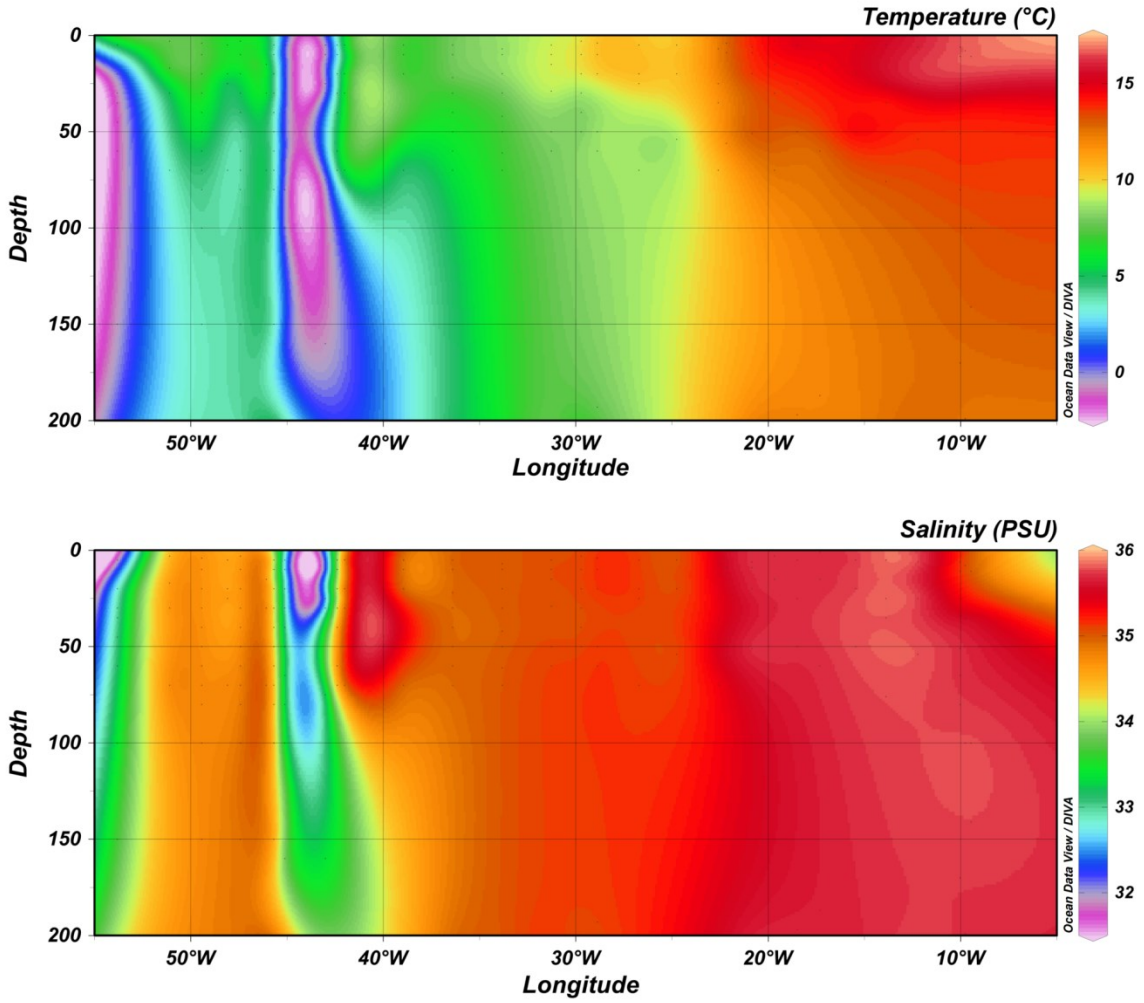
Supplemental Figure 13: RNA transcript of *nifH*.

The isolated diazotroph was grown on nitrogen depleted artificial sea water. RNA was extracted once cultures had grown to visible density. The RNA was reverse transcripted into cDNA and *nifH* PCR was performed (Zehr et al. 2001). RNA and negative controls were included in the reaction.



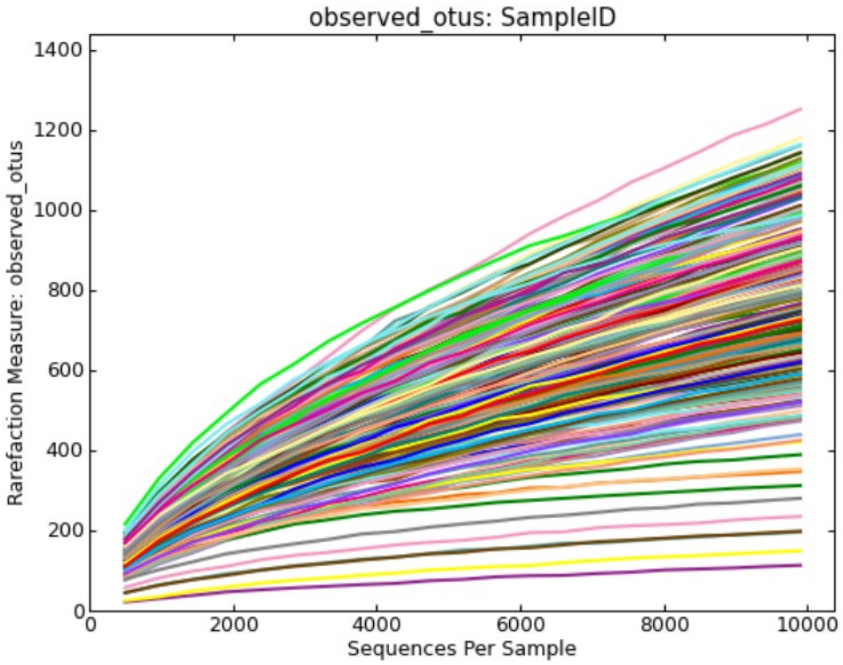
Supplemental Figure 14: Environmental conditions and the isolate's abundances throughout the North Atlantic Ocean.

(A) depicts the dendogram of the Cluster Analysis from the environmental matrix. Samples were clustered based on their Euclidian-distance. Significant clustering occurred on black lines; red dotted lines are non-significant. Samples are colour coded according to region as indicated in the figure legend. (B) shows the PCA analysis of the same samples. Sample origin is labelled according to the legend. (C) highlights samples with especially high abundances in the GEOVIDE cruise transect and the Bedford Basin.



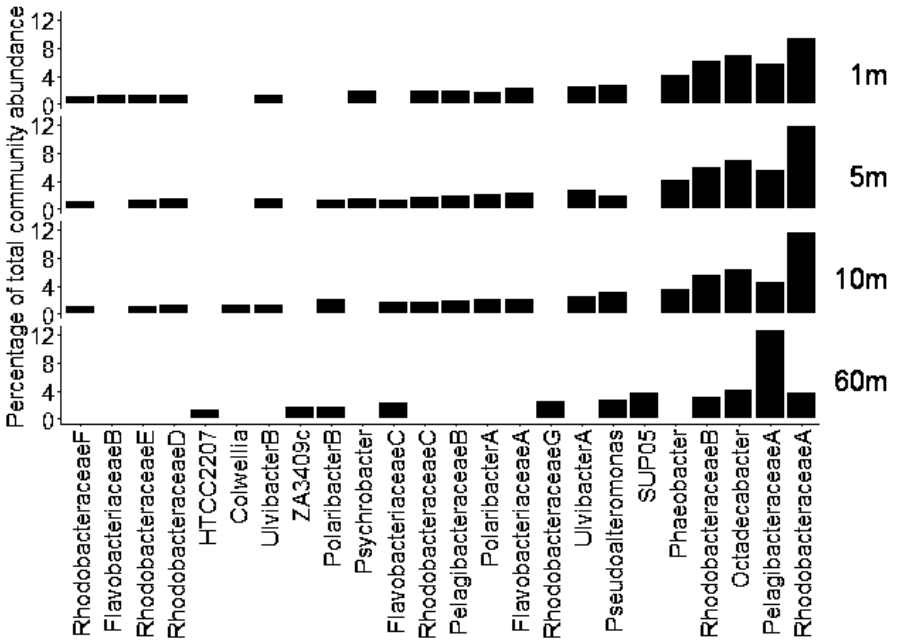
Supplemental Figure 15: Temperature and salinity measurements during the GEOVIDE cruise.

Temperature (°C) and salinity (PSU) measurements as they were taken during CTD casts over the GEOVIDE cruise transect.



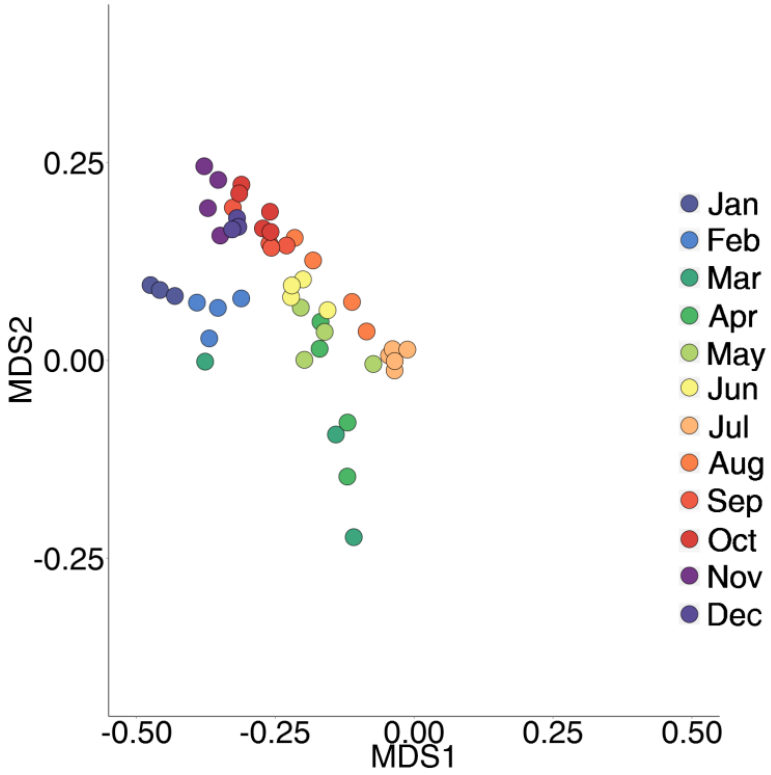
Supplemental Figure 16: Alpha-Rarefaction curve of 16S rRNA Bedford Basin samples collected in 2014.

Displayed is the number of observed OTUs at a specific number of sequences in one sample up to the chosen 9,900 representative sequences.



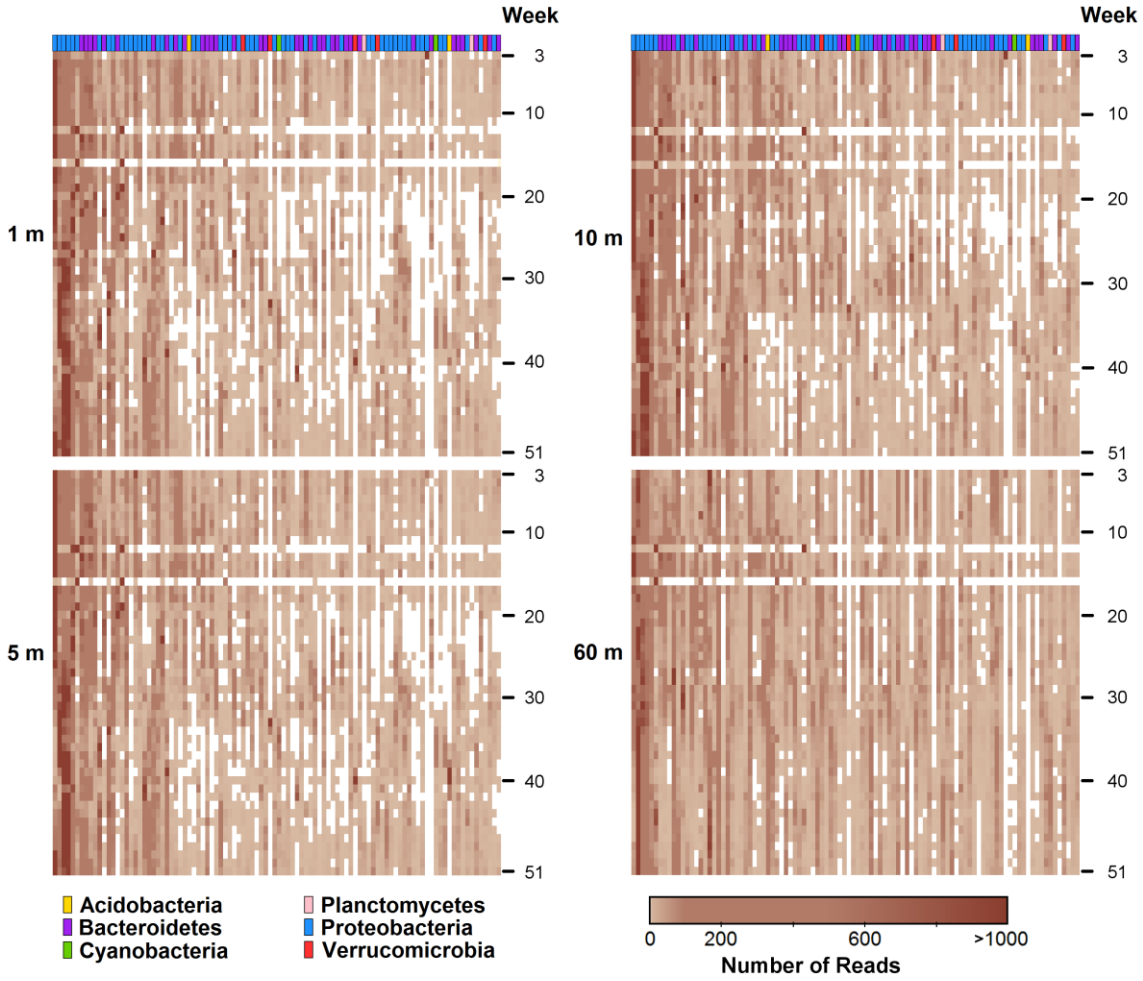
Supplemental Figure 17: Relative abundances of V6-V8 16S rRNA OTUs making up more than 1% of all reads in the Bedford Basin.

Depicted are the relative abundances of the OTUs that made up more than 1% of all reads at each depth.



Supplemental Figure 18: Community structure of weekly samples at 60 m depth in the Bedford Basin in 2014.

Non-linear Multi-Dimensional (NMDS) was used to plot sample similarity according to their taxonomic composition and abundance. Species counts were Hellinger transformed and NMDS plots were created based on Bray-Curtis dissimilarity between samples using *vegan* in R (R Core Team 2015; Dixon 2003). Each point represents one sample and is colour coded according to month.



Supplemental Figure 19: Occurrences of the 100 most common OTUs in the Bedford Basin in 2014.

Number of reads each of the 100 most common OTUs displayed at 1, 5, 10 and 60 m depth from week 3 to week 51 in 2014. Taxonomic classification of each OTU at the phylum level is indicated according to the colour legend.

APPENDIX C: COPYRIGHT PERMISSION

Elsevier does not required copyright permissions for publications included in theses.

The permissions website (https://www.elsevier.com/about/company-information/policies/copyright/permissions#Permission%20Guidelines, March 28, 2017) states:

Can I include/use my article in my thesis/dissertation?

Yes. Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.